



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 121628

TO: Jeffrey Parkin
Location: rem/3c18
Art Unit: 1648
Thursday, May 13, 2004

Case Serial Number: 09/549186

From: Alex Waclawiw
Location: Biotech-Chem Library
Rem 1A71
Phone: 272-2534

Alexandra.waclawiw@uspto.gov

Search Notes

This Page Blank (uspto)

SEARCH REQUEST FORM

Access DB# 121628

Scientific and Technical Information Center

Requester's Full Name: PARKER, JEFF Examiner #: 22607 Date: 05/07/04
 Art Unit: 1648 Phone Number: 2-0908 Serial Number: 09/549,186
 Mail Box and Bldg Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need. BOTH ME'

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept on which the invention is based. Define any terms that may have a special meaning. Give examples of relevant citations, authors, etc. in the field. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

For Sequence Searches Only Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.

Please search the following RETROINVERSO peptides:

1) X[L-Q-R-A-V-R-P-A-L-S-G-F-D-G-R-V-G-S-G-C]Z
 (INVERSE OF SEQ 7)

2) X[L-Q-R-A-U-R-L-A-L-S-G-F-D-G-R-U-G-S-G-C]Z
 (INVERSE OF SEQ 8)

3) X[L-Q-R-A-U-R-L-A-L-S-G-S-D-G-R-U-G-S-G-C]Z
 (INVERSE OF SEQ 9)

STAFF USE ONLY

Point of Contact: Alexandra Wacławski
 Searcher: Technical Info. Specialist
 Searcher Phone #: 211-5602 Tel: 300-4491
 Searcher Location: _____

Type of Search

NA Sequence (#) _____

AA Sequence (#) 3

Structure (#) _____

Vendors and cost where applicable

STN _____

Dialog _____

Questel/Orbit _____

Date Searcher Picked Up

5-10-04

STN
 @
 S

This Page Blank (uspto)

=> d his 18-

(FILE 'REGISTRY' ENTERED AT 12:59:48 ON 10 MAY 2004)

FILE 'REGISTRY' ENTERED AT 13:02:58 ON 10 MAY 2004

L8 0 S LQRAVR[PL]ALSG[FS]DGRVGSGC/SQSP
L9 2 S CGSGVRGD[FS]GSLA[LP]RVARQL/SQSP

FILE 'REGISTRY' ENTERED AT 13:05:14 ON 10 MAY 2004

FILE 'CAPLUS' ENTERED AT 13:05:16 ON 10 MAY 2004

L10 3 S L9

=> fil reg

FILE 'REGISTRY' ENTERED AT 13:06:17 ON 10 MAY 2004

USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.

PLEASE SEE "HELP USAGETERMS" FOR DETAILS.

COPYRIGHT (C) 2004 American Chemical Society (ACS)

Property values tagged with IC are from the ZIC/VINITI data file provided by InfoChem.

STRUCTURE FILE UPDATES: 9 MAY 2004 HIGHEST RN 680971-82-8

DICTIONARY FILE UPDATES: 9 MAY 2004 HIGHEST RN 680971-82-8

TSCA INFORMATION NOW CURRENT THROUGH JANUARY 6, 2004

Please note that search-term pricing does apply when conducting SmartSELECT searches.

Crossover limits have been increased. See HELP CROSSOVER for details.

Experimental and calculated property data are now available. For more information enter HELP PROP at an arrow prompt in the file or refer to the file summary sheet on the web at:

<http://www.cas.org/ONLINE/DBSS/registryss.html>

=> d que 18

L8 0 SEA FILE=REGISTRY ABB=ON PLU=ON LQRAVR[PL]ALSG[FS]DGRVGSGC/SQ
SP

=> d que 19

L9 2 SEA FILE=REGISTRY ABB=ON PLU=ON CGSGVRGD[FS]GSLA[LP]RVARQL/SQ
SP

=> d sqide 1-2 19

L9 ANSWER 1 OF 2 REGISTRY COPYRIGHT 2004 ACS on STN

RN 199790-86-8 REGISTRY

CN L-Leucine, N-acetyl-L-cysteinylglycyl-L-serylglycyl-L-valyl-L-arginylglycyl-L- α -aspartyl-L-phenylalanylglycyl-L-seryl-L-leucyl-L-alanyl-L-prolyl-L-arginyl-L-valyl-L-alanyl-L-arginyl-L-glutaminy- (9CI)
(CA INDEX NAME)

FS PROTEIN SEQUENCE; STEREOSEARCH

SQL 20

NTE modified

type location description

This Page Blank (uspto)

terminal mod. Cys-1 - N-acetyl

SEQ 1 CGSGVRGDFG SLAPRVARQL

HITS AT: 1-20

RELATED SEQUENCES AVAILABLE WITH SEQLINK

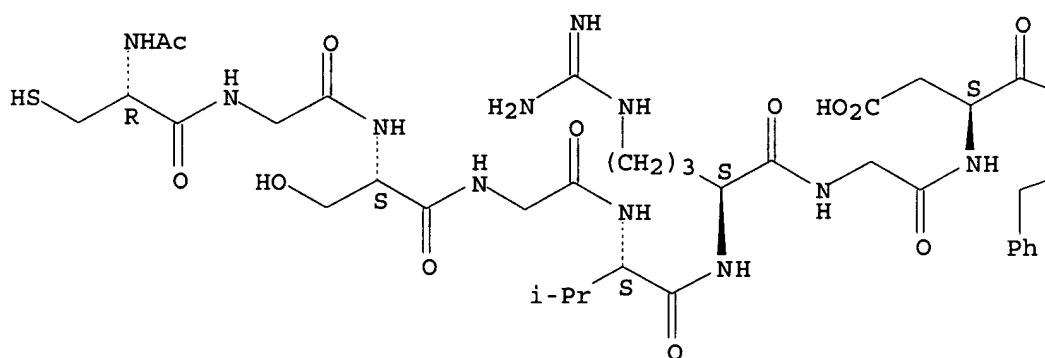
MF C88 H146 N30 O27 S

SR CA

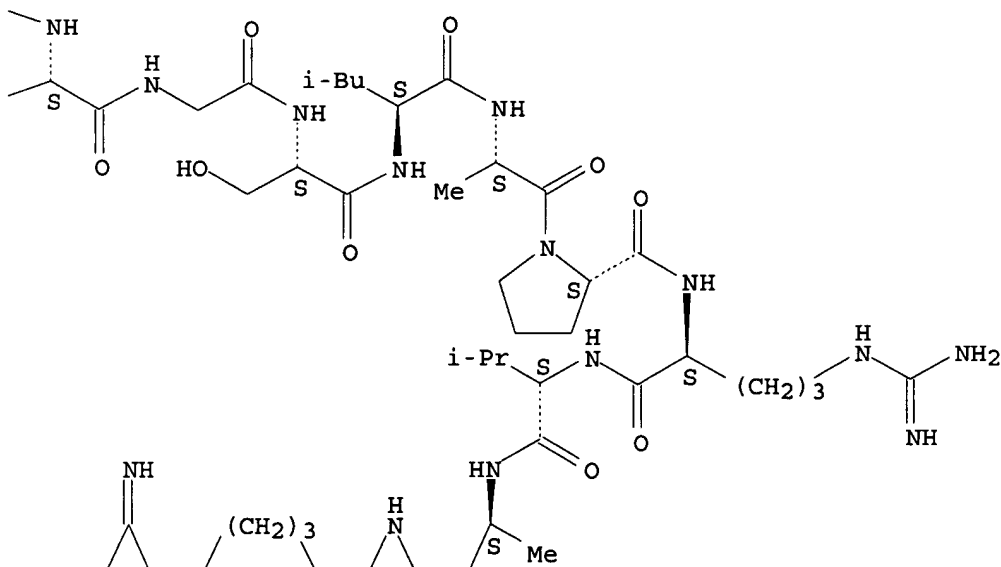
LC STN Files: CA, CAPLUS, TOXCENTER

Absolute stereochemistry.

PAGE 1-A

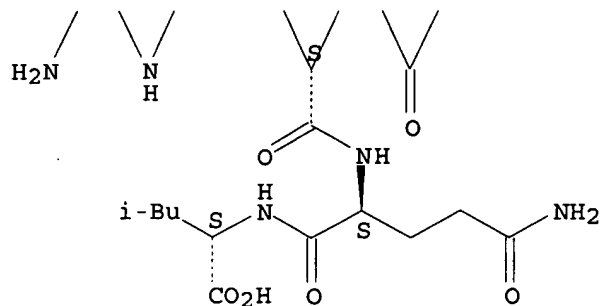


PAGE 1-B



This Page Blank (uspto)

PAGE 2-B



3 REFERENCES IN FILE CA (1907 TO DATE)
 1 REFERENCES TO NON-SPECIFIC DERIVATIVES IN FILE CA
 3 REFERENCES IN FILE CAPLUS (1907 TO DATE)

L9 ANSWER 2 OF 2 REGISTRY COPYRIGHT 2004 ACS on STN
 RN 199790-81-3 REGISTRY
 CN L-Leucine, L-cysteinylglycyl-L-serylglycyl-L-valyl-L-arginylglycyl-L-
 α-aspartyl-L-phenylalanylglycyl-L-seryl-L-leucyl-L-alanyl-L-prolyl-L-
 arginyl-L-valyl-L-alanyl-L-arginyl-L-glutaminy- (9CI) (CA INDEX NAME)
 FS PROTEIN SEQUENCE; STEREOSEARCH
 SQL 20

SEQ 1 CGSGVRGDFG SLAPRVARQL
 =====

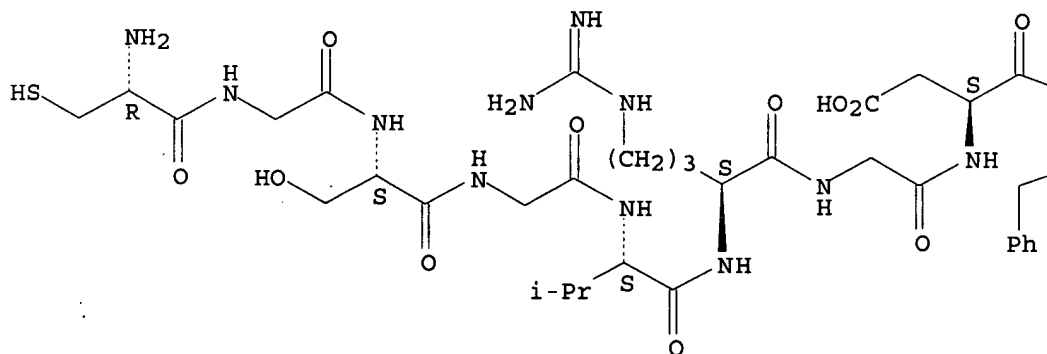
HITS AT: 1-20

RELATED SEQUENCES AVAILABLE WITH SEQLINK

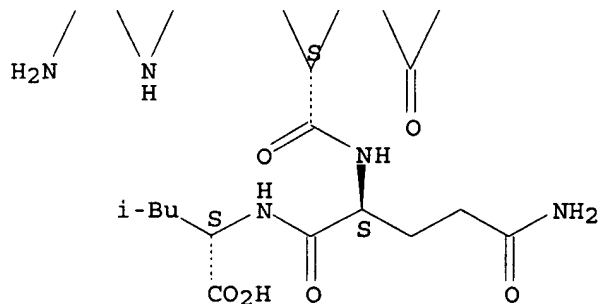
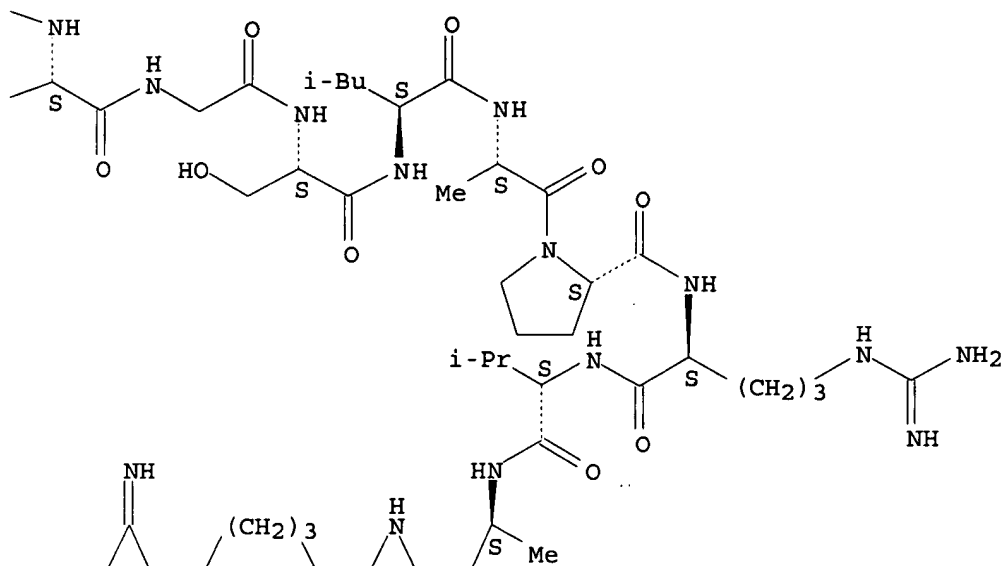
MF C86 H144 N30 O26 S
 SR CA
 LC STN Files: CA, CAPLUS

Absolute stereochemistry.

PAGE 1-A



This Page Blank (uspto)



2 REFERENCES IN FILE CA (1907 TO DATE)
2 REFERENCES IN FILE CAPLUS (1907 TO DATE)

=> fil hcaplus
FILE 'HCAPLUS' ENTERED AT 13:06:43 ON 10 MAY 2004
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.
COPYRIGHT (C) 2004 AMERICAN CHEMICAL SOCIETY (ACS)

Copyright of the articles to which records in this database refer is held by the publishers listed in the PUBLISHER (PB) field (available for records published or updated in Chemical Abstracts after December 26, 1996), unless otherwise indicated in the original publications. The CA Lexicon is the copyrighted intellectual property of the

This Page Blank (uspto)

the American Chemical Society and is provided to assist you in searching databases on STN. Any dissemination, distribution, copying, or storing of this information, without the prior written consent of CAS, is strictly prohibited.

FILE COVERS 1907 - 10 May 2004 VOL 140 ISS 20
FILE LAST UPDATED: 9 May 2004 (20040509/ED)

This file contains CAS Registry Numbers for easy and accurate substance identification.

'OBI' IS DEFAULT SEARCH FIELD FOR 'HCAPLUS' FILE

=> fil caplus
FILE 'CAPLUS' ENTERED AT 13:06:52 ON 10 MAY 2004
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.
COPYRIGHT (C) 2004 AMERICAN CHEMICAL SOCIETY (ACS)

Copyright of the articles to which records in this database refer is held by the publishers listed in the PUBLISHER (PB) field (available for records published or updated in Chemical Abstracts after December 26, 1996), unless otherwise indicated in the original publications. The CA Lexicon is the copyrighted intellectual property of the American Chemical Society and is provided to assist you in searching databases on STN. Any dissemination, distribution, copying, or storing of this information, without the prior written consent of CAS, is strictly prohibited.

FILE COVERS 1907 - 10 May 2004 VOL 140 ISS 20
FILE LAST UPDATED: 9 May 2004 (20040509/ED)

This file contains CAS Registry Numbers for easy and accurate substance identification.

'OBI' IS DEFAULT SEARCH FIELD FOR 'CAPLUS' FILE

=> d que nos l10
L9 2 SEA FILE=REGISTRY ABB=ON PLU=ON CGSGVRGD[FS]GSLA[LP]RVARQL/SQ
SP
L10 3 SEA FILE=CAPLUS ABB=ON PLU=ON L9

=> d .ca l10 1-5

L10 ANSWER 1 OF 3 CAPLUS COPYRIGHT 2004 ACS on STN
ACCESSION NUMBER: 2003:489994 CAPLUS
DOCUMENT NUMBER: 139:83614
TITLE: Intranasal immunization of guinea pigs with an immunodominant foot-and-mouth disease virus peptide conjugate induces mucosal and humoral antibodies and protection against challenge
AUTHOR(S): Fischer, D.; Rood, D.; Barrette, R. W.; Zuwallack, A.; Kramer, E.; Brown, F.; Silbart, L. K.
CORPORATE SOURCE: Center of Excellence for Vaccine Research, University of Connecticut, Storrs, CT, USA
SOURCE: Journal of Virology (2003), 77(13), 7486-7491
CODEN: JOVIAM; ISSN: 0022-538X
PUBLISHER: American Society for Microbiology

This Page Blank (uspto)

DOCUMENT TYPE: Journal
LANGUAGE: English

AB Guinea pigs immunized intranasally with a keyhole limpet hemocyanin-linked peptide, corresponding to the prominent G-H loop of the VP1 protein of foot-and-mouth disease virus, raised substantial levels of antipeptide and virus-neutralizing antibodies in sera and of peptide-specific secretory IgA in nasal secretions. In groups of animals immunized intranasally without adjuvant, 86 percent were fully protected upon challenge with homotypic virus. Surprisingly, animals given the peptide conjugates plus the mucosal adjuvant cholera toxin were afforded only partial protection in that primary lesions were observed in most animals, although spread to other feet was prevented. These results indicate that intranasal inoculation with the peptide offers a potential route of vaccination against foot-and-mouth disease and may be useful for eliciting protection in the upper respiratory tracts of susceptible animals.

CC 15-2 (Immunochemistry)

IT 199790-86-8D, conjugates with carrier protein
RL: PAC (Pharmacological activity); BIOL (Biological study)
(intranasal immunization of guinea pigs with an immunodominant foot-and-mouth disease virus peptide conjugate induces mucosal and humoral antibodies and immune protection)

REFERENCE COUNT: 42 THERE ARE 42 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L10 ANSWER 2 OF 3 CAPLUS COPYRIGHT 2004 ACS on STN

ACCESSION NUMBER: 1998:597752 CAPLUS

DOCUMENT NUMBER: 129:342490

TITLE: Mimicry of an immunodominant epitope of foot and mouth disease virus with retro-inverso isomers: a new approach in the design of peptide based vaccines

AUTHOR(S): Guichard, Gilles; Benkirane, Nadia; Briand, Jean-Paul; Muller, Sylviane; Van Regenmortel, Marc H. V.; Newman, John F. E.; Brown, Fred

CORPORATE SOURCE: Institut de Biologie Moleculaire et Cellulaire, UPR 9021 CNRS, Strasbourg, 67000, Fr.

SOURCE: Peptides 1996, Proceedings of the European Peptide Symposium, 24th, Edinburgh, Sept. 8-13, 1996 (1998), Meeting Date 1996, 447-448. Editor(s): Ramage, Robert; Epton, Roger. Mayflower Scientific: Kingswinford, UK.
CODEN: 66RCA5

DOCUMENT TYPE: Conference

LANGUAGE: English

AB The retro-inverso analog of the immunodominant epitope 141-159 of the VP1 protein of foot-and-mouth disease virus was shown to induce in rabbits a peptide-specific neutralizing IgG response of much higher titer which lasted longer than that induced by the epitope.

CC 15-2 (Immunochemistry)

IT 199790-79-9 199790-81-3 199790-86-8 215603-35-3

RL: BAC (Biological activity or effector, except adverse); BSU (Biological study, unclassified); THU (Therapeutic use); BIOL (Biological study); USES (Uses)

(mimicry of an immunodominant epitope of foot and mouth disease virus with retro-inverso isomers and induction of neutralizing antibodies)

REFERENCE COUNT: 4 THERE ARE 4 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

L10 ANSWER 3 OF 3 CAPLUS COPYRIGHT 2004 ACS on STN

ACCESSION NUMBER: 1997:768652 CAPLUS

DOCUMENT NUMBER: 128:33517

This Page Blank (uspto)

TITLE: A retro-inverso peptide corresponding to the GH loop of foot-and-mouth disease virus elicits high levels of long-lasting protective neutralizing antibodies

AUTHOR(S): Briand, Jean-Paul; Benkirane, Nadia; Guichard, Gilles; Newman, John F. E.; Van Regenmortel, Marc H. V.; Brown, Fred; Muller, Sylviane

CORPORATE SOURCE: Unite Propre de Recherche 9021, Centre National de la Recherche Scientifique, Institut de Biologie Moleculaire et Cellulaire, Strasbourg, 67084, Fr.

SOURCE: Proceedings of the National Academy of Sciences of the United States of America (1997), 94(23), 12545-12550
CODEN: PNASA6; ISSN: 0027-8424

PUBLISHER: National Academy of Sciences

DOCUMENT TYPE: Journal

LANGUAGE: English

AB Peptides corresponding to the immunodominant loop located at residues 135-158 on capsid protein VP1 of foot-and-mouth disease virus (FMDV) generally elicit high levels of anti-peptide and virus-neutralizing antibodies. In some instances, however, the level of neutralizing antibodies is low or even negligible, even though the level of anti-peptide antibodies is high. The authors have shown previously that the antigenic activity of peptide 141-159 of VP1 of a variant of serotype A can be mimicked by a retro-inverso (all-D retro or retroenantio) peptide analog. This retro-inverso analog induced greater and longer-lasting antibody titers than did the corresponding L-peptide. The authors now show that a single inoculation of the retro-inverso analog elicits high levels of neutralizing antibodies that persist longer than those induced against the corresponding L-peptide and confer substantial protection in guinea pigs challenged with the cognate virus. In view of the high stability to proteases of retro-inverso peptide analogs and their enhanced immunogenicity, these results have practical relevance in designing potential peptide vaccines.

CC 15-2 (Immunochemistry)

IT 164259-71-6 199790-79-9 199790-81-3 199790-82-4
199790-83-5 199790-84-6 199790-86-8 199790-89-1
199790-90-4

RL: BAC (Biological activity or effector, except adverse); BSU (Biological study, unclassified); PRP (Properties); BIOL (Biological study)
(retro-inverso peptide corresponding to GH loop of foot-and-mouth disease virus elicitation of protective neutralizing antibodies in relation to vaccine)

REFERENCE COUNT: 30 THERE ARE 30 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

=>

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2004, 06:49:47 ; Search time 48 Seconds
(without alignments)
117.728 Million cell updates/sec

Title: 09549186-7

Perfect score: 104

Sequence: 1 LQAVRPALSGDGRVSGC 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : A_Geneseq_29Jan04:*

1: Geneseq1980s:*

2: Geneseq1990s:*

3: Geneseq2000s:*

4: Geneseq2001s:*

5: Geneseq2002s:*

6: Geneseq2003as:*

7: Geneseq2003bs:*

8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	88	84.6	20	2	ADB87347 Vaccinati
2	86	82.7	19	2	AAR45910 Foot and
3	49	47.1	785	4	ABB61535 Drosophil
4	46	44.2	339	7	ADC94164 E. faeciu
5	45.5	43.8	229	4	AAG92858 C. glutami
6	45	43.3	67	4	AUA43462 Propionib
7	45	43.3	67	6	ABM39981 Propionib
8	45	43.3	141	4	AUA63931 Propionib
9	45	43.3	141	6	ABM60450 Propionib
10	45	43.3	172	4	AAB90645 Human sec
11	45	43.3	287	6	ABU48621 Protein e
12	45	43.3	484	4	AUA34790 E. coli c
13	45	43.3	484	6	ABU28815 Protein e
14	44.5	42.8	73	5	ABP34437 Human ORF
15	44.5	42.8	291	6	ABP76931 N. gonorr
16	44	42.3	132	4	ABG05276 Novel hum
17	44	42.3	943	5	ABG91056 Neisseria
18	44	42.3	1726	2	AAW38756 Phosphati
19	44	42.3	1964	2	AAW95557 Mus muscu
20	43.5	41.8	192	6	ABM69770 Photorhab
21	43.5	41.8	290	6	ABP76938 N. gonorr
22	43.5	41.8	481	6	ABP76945 N. gonorr
23	43.5	41.8	498	2	AAW38584 Neisseria
24	43.5	41.8	498	3	AAW74707 Neisseria
25	43	41.3	97	6	ABM65962 Propionib

26	43	41.3	228	2	AAR98742	Aar98742 Tazaroten
27	43	41.3	445	6	ABR41629	Abr41629 Human DIT
28	43	41.3	961	6	ABR41625	Abr41625 Human DIT
29	43	41.3	1121	4	AAG78887	Aag78887 Human fib
30	43	41.3	1169	4	ABG68757	Abg68757 Drosophil
31	43	41.3	2895	5	ABG94632	Abg94632 Human NOV
32	43	41.3	2757	5	ABG94633	Abg94633 Human NOV
33	43	41.3	2809	5	AAG66169	Aag66169 Human fib
34	43	41.3	2844	5	ABG94629	Abg94629 Human NOV
35	43	41.3	2845	5	ABG94631	Abg94631 Human NOV
36	43	41.3	2877	5	ABG94630	Abg94630 Human NOV
37	42.5	40.9	3640	4	ABG23029	Abg23029 Novel hum
38	42	40.4	59	4	AAU40754	Aau40754 Propionib
39	42	40.4	59	6	ABM37273	Abm37273 Propionib
40	42	40.4	79	4	AAU47704	Aau47704 Propionib
41	42	40.4	79	6	ABM44223	Abm44223 Propionib
42	42	40.4	97	4	AAU48847	Aau48847 Propionib
43	42	40.4	97	6	ABM45366	Abm45366 Propionib
44	42	40.4	99	7	ADB74327	Adb74327 Mycobacte
45	42	40.4	115	7	ADB74546	Adb74546 Mycobacte
46	42	40.4	117	4	AAU40418	Aau40418 Propionib
47	42	40.4	117	6	ABM36937	Abm36937 Propionib
48	42	40.4	155	6	ADA36132	Ada36132 Acinetoba
49	42	40.4	174	4	ABM69946	Abm69946 Drosophil
50	42	40.4	263	5	ABB48000	Abb48000 Listeria
51	42	40.4	263	6	ABU32866	Abu32866 Protein e
52	42	40.4	273	6	ABU21562	Abu21562 Protein e
53	42	40.4	301	4	ABM69239	Abm69239 Drosophil
54	42	40.4	466	3	AAW44745	Aay44745 Soybean A
55	42	40.4	466	5	ABM92610	Abm92610 Herbicida
56	42	40.4	551	6	ABU21914	Abu21914 Protein e
57	42	40.4	799	5	ABP65781	Abp65781 Bifidobac
58	42	40.4	812	2	AAR12934	Aar12934 Plasminog
59	42	40.4	898	2	AAW89413	Aaw89413 Moraxella
60	41.5	39.9	101	2	AAR81440	Aar81440 Hepatitis
61	41.5	39.9	101	3	ABM90917	Abm90917 Hepatitis
62	41.5	39.9	283	5	ABM92103	Abm92103 Herbicida
63	41.5	39.9	343	5	ABM72332	Abm72332 Rat prote
64	41.5	39.9	385	3	AAW74709	Aay74709 Neisseria
65	41.5	39.9	396	6	ABU48676	Abu48676 Protein e
66	41.5	39.9	482	2	AAW38582	Aay38582 Neisseria
67	41.5	39.9	483	2	AAW38581	Aay38581 Neisseria
68	41.5	39.9	483	3	AAW74708	Aay74708 Neisseria
69	41.5	39.9	984	5	AAE19794	Aae19794 Human Hai
70	41.5	39.9	984	6	ABU62540	Abu62540 Human Hai
71	41.5	39.9	984	6	ABU62545	Abu62545 Human Hai
72	41.5	39.9	986	6	ABU14884	Abu14884 Protein e
73	41.5	39.9	1189	2	AAW15217	Aay15217 Human Hai
74	41.5	39.9	1189	2	AAW15218	Aay15218 Human Hai
75	41.5	39.9	1189	5	AAE19795	Aae19795 Human put
76	41.5	39.9	1189	5	AAE19796	Aae19796 Human Hai
77	41.5	39.9	1189	6	ABG72775	Abg72775 Human hai
78	41.5	39.9	1189	6	ABG62542	Abg62542 Human Hai
79	41.5	39.9	1189	6	ABU62541	Abu62541 Human Hai
80	41	39.4	73	2	AAW22514	Aay22514 Xenopus h
81	41	39.4	118	4	AAU49158	Aau49158 Propionib
82	41	39.4	118	6	ABM45677	Abm45677 Propionib
83	41	39.4	134	2	AAW90780	Aar90780 HCV antig
84	41	39.4	138	4	AAW56075	Aus56075 Propionib
85	41	39.4	138	6	ABM52594	Abm52594 Propionib
86	41	39.4	208	7	ADC87091	Adc87091 Human GPC
87	41	39.4	220	4	ABG29258	Abg29258 Novel hum
88	41	39.4	231	6	ABR41590	Abr41590 Human DIT
89	41	39.4	339	2	AAW81592	Aaw81592 Protein e
90	41	39.4	403	2	AAW81588	Aaw81588 Protein e
91	41	39.4	484	6	ABU47531	Abu47531 Protein e
92	41	39.4	512	6	ABU21483	Abu21483 Protein e
93	41	39.4	597	3	AAW90257	Aay90257 Streptoco
94	41	39.4	850	2	AAW62040	Aaw62040 Protein i
95	41	39.4	894	2	AAW89417	Aaw89417 Moraxella
96	41	39.4	1064	7	ADB65676	Adb65676 Human pro
97	40.5	38.9	50	4	AAU46350	Aau46350 Propionib
98	40.5	38.9	50	6	ABM42869	Abm42869 Propionib

99 40.5 38.9 53 4 AAU44797 Propionib
100 40.5 38.9 53 6 ABM41316 Propionib

ALIGNMENTS

RESULT 1
ADB87347
ID ADB87347 standard; peptide; 20 AA.
XX
AC ADB87347;
XX
DT 04-DEC-2003 (first entry)
XX
DE Vaccination related retro-partly Inverso peptide #1.
XX
XX immunoretrold; anti-immunoretrold; CONH linkage; NHCO linkage;
KW retropeptide; retroinverso peptide; vaccine; viral; bacterial infection;
KW autoimmune disease; neurodegenerative disease; retro-partly;
KW Inverso peptide.
XX
OS Unidentified.
FH Key Location/Qualifiers
FT Modified-site 1 /note= "Modified by OH-m. Peptide can be an R or S
FT isomer"
FT Misc-difference 2 /note= "D-form residue"
FT Misc-difference 3 /note= "D-form residue"
FT Misc-difference 4 /note= "D-form residue"
FT Misc-difference 5 /note= "D-form residue"
FT Misc-difference 6 /note= "D-form residue"
FT Misc-difference 7 /note= "D-form residue"
FT Misc-difference 8 /note= "D-form residue"
FT Misc-difference 9 /note= "D-form residue"
FT Misc-difference 10 /note= "D-form residue"
FT Misc-difference 11 /note= "D-form residue"
FT Misc-difference 12 /note= "D-form residue"
FT Misc-difference 13 /note= "D-form residue"
FT Misc-difference 14 /note= "D-form residue"
FT Misc-difference 15 /note= "D-form residue"
FT Misc-difference 16 /note= "D-form residue"
FT Misc-difference 17 /note= "D-form residue"
FT Modified-site 18 /note= "C-terminal amide"
XX FR2717081-A1.
XX
XX 15-SEP-1995.
XX
XX 14-MAR-1994; 94FR-00002950.
XX
XX 14-MAR-1994; 94FR-00002950.
XX
XX (CNRS) CENT NAT RECH SCI.
XX
XX Guichard G, Muller S, Briand J, Regenmortel MHV;
XX

DR WPI; 1995-322414/42.
XX
PT Therapeutic and diagnostic uses of retro peptide analogues - corresp. to
PT parent peptide chains with CONH linkages replaced by NHCO linkages, also
XX antibodies against the peptide(s).
PS Disclosure; Page 21; 58pp; French.
XX
XX This invention relates to the novel uses of 'immunoretrold' or anti-
CC immunoretrold antibodies, where the immunoretrold is peptide analogues
CC in which one or more (preferably all) of the CONH linkages in the chain
CC of the corresponding parent peptides are replaced by NHCO linkages and
CC the chirality of each amino acid residue, whether involved in NHCO
CC linkages or not, is either conserved or inverted with regards to the
CC corresponding amino acid residue in the parent peptides. For example,
CC 'retropeptides' or 'retroinverso peptides', provided that the
CC immunoretrold is capable of forming complexes with the anti-
CC immunoretrold antibodies and with antibodies directed against the parent
CC peptides or parent proteins and/or the parent peptide enantiomers or
CC parent protein enantiomers. The immunoretrold is used to prepare
CC medicaments for preventing or treating pathologies associated with the
CC presence of an exogenous or endogenous protein capable of being
CC implicated directly or indirectly in the appearance and/or development of
CC the pathologies. Immunoretrold can also be used to prepare vaccines for
CC preventing pathologies associated with the presence of an exogenous or
CC endogenous protein recognised by antibodies directed against
CC immunoretrold. Comparisons containing immunoretrold associated with a
CC carrier molecule capable of inducing production of antibodies against an
CC exogenous or endogenous protein responsible for a pathology, or of
CC inducing a cytotoxic cellular immune response are useful as vaccines.
CC Pathologies that can be diagnosed or treated are especially viral or
CC bacterial infections, autoimmune diseases and neurodegenerative diseases.
CC This sequence represents a vaccination related retro-partly Inverso
XX peptide relating to the retropeptides of the invention.
SQ Sequence 20 AA;
Query Match 84.6%; Score 88; DB 2; Length 20;
Best Local Similarity 90.0%; Pred. No. 5 6e-07;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 LQRAVRPALSGDGRVSGC 20
DB 1 LQRAVRXALSGXDRVSGC 20
RESULT 2
AAR45910
ID AAR45910 standard; peptide; 19 AA.
XX
AC AAR45910;
XX
XX 27-AUG-2003 (revised)
DT 25-MAR-2003 (revised)
DT 27-JUL-1994 (first entry)
XX
DE Foot and Mouth Disease Virus immunogenic peptide P6.
XX
KW immunodominant surface-exposed epitope; T-cell proliferation;
KW stimulation; immunogenicity; increase; enhance; vaccine;
KW Foot and Mouth Disease Virus; FMDV; Aphthovirus; LCP; lipid anchor;
KW immunogenic carrier.
XX
OS Foot-and-mouth disease virus.
XX
XX WO9402506-A1.
XX
XX 03-FEB-1994.
PD
XX 23-JUL-1993; 93WO-CB001558.
PF
XX 24-JUL-1992; 92GB-00015780.
PR
XX

PA (UNLO) UNIV LONDON SCHOOL PHARMACY.
XX Toth I, Gibbons WA;
XX WPI; 1994-048791/06.
XX New lipidic amino acid based anchor system - for attachment of short
PT synthetic peptide(s) to enhance their antigenicity and for use, e.g., as
PT vaccines.
XX
XX Example 3; Page 36; 50pp; English.
XX A lipidic amino acid based anchor system was synthesised. Eight copies of
CC the FMDV immunogenic peptide P6 (AAR45910) were attached to the anchor to
CC give (peptide)8LYe4LYs2LYs(HNCH((CH2)13Me)CO)3NH2. When injected into
CC cows, immunogenicity was found to be 10 times higher than would be
CC expected from a conventional BSA- or KLH-conjugated FMDV. (Updated on 25-
CC MAR-2003 to correct PN field.) (Updated on 27-AUG-2003 to correct OS
CC field.)
XX
XX Sequence 19 AA;
SQ

Query Match 82.7%; Score 86; DB 2; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.1e-06;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RAVRPALSGFDRGVSG 19
|||
DB 3 RAVRPALSGFDRGVSG 19
|||

RESULT 3
ABB61535
ID ABB61535 standard; protein; 785 AA.
XX
AC ABB61535;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 11397.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
XX Drosophila melanogaster.
OS
XX WO200171042-A2.
PN
XX
PD 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
PF
XX
XX 23-MAR-2000; 2000US-0191637P.
PR
XX 11-JUL-2000; 2000US-00614150.
PR
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX WPI; 2001-656860/75.
DR N-PSDB; ABL05638.
DR
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
XX Disclosure; SEQ ID NO 11397; 21pp + Sequence Listing; English.
PS
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 785 AA;
SQ

Query Match 47.1%; Score 49; DB 4; Length 785;
Best Local Similarity 60.0%; Pred. No. 49;
Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 6 RPALSGFDRGVSGC 20
|||
DB 173 RPTLSGFGVASEGC 187
|||

RESULT 4
ADC94164
ID ADC94164 standard; protein; 339 AA.
XX
AC ADC94164;
XX
XX 01-JAN-2004 (first entry)
DT
XX
XX E. faecium protein sequence SEQ ID 3791.
DE
XX Vaccine; urinary tract infection; bacteraemia; endocarditis; wound;
KW abdominal-pelvic infection.
KW
XX Enterococcus faecium.
OS
XX US6583275-B1.
PN
XX 24-JUN-2003.
PD
XX
XX 30-JUN-1998; 98US-00107532.
PF
XX
XX 02-JUL-1997; 97US-0051571P.
PR
XX 14-MAY-1998; 98US-0085598P.
PR
XX (GENO-) GENOME THERAPEUTICS CORP.
PA
XX
XX Doucette-Stamm LA, Bush D;
PI
XX
XX WPI; 2003-799836/75.
DR N-PSDB; ADC90510.
DR
XX
XX New isolated nucleic acid derived from Enterococcus faecium encoding an
PT Enterococcus faecium polypeptide useful for detection, prevention and
PT treatment of a pathological condition resulting from a bacterial
PT infection.
XX
XX Example 1; SEQ ID NO 3791; 243pp; English.
PS
XX The invention relates to an isolated nucleic acid derived from
CC Enterococcus faecium encoding an Enterococcus faecium polypeptide having
CC one of 10 fully defined sequences given in the (or comprising 40
CC sequential nucleotides chosen from any of the nucleic acids, its
CC complement or sequences hybridising to it). Also included are a
CC recombinant vector comprising the nucleic acid operably linked to
CC transcription regulatory element, a cell comprising the vector and a
CC single-stranded probe comprising the nucleic acid. The nucleic acids are
CC chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.
CC The nucleic acids are useful for diagnosing pathological conditions
CC resulting from E. faecium bacterial infection (e.g. urinary tract
CC infection, bacteraemia, endocarditis, wounds and abdominal-pelvic
CC infection) and for screening drugs such as agonists and antagonists. The
CC nucleic acid is useful for recombinant production of Candida albicans -
CC derived peptides or antisense polypeptides. Pharmaceutical compositions
CC and vaccines containing the nucleic acid are useful for preventing or
CC treating Enterococcus faecium infections. The present sequence represents
CC one if the disclosed E. faecium proteins.

```
XX SQ Sequence 339 AA;
Query Match 44.2%; Score 46; DB 7; Length 339;
Best Local Similarity 57.9%; Pred. No. 62;
Matches 11; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 LQRAVRPALSGFDGRVSG 19
   ||: ||| ||| ||| |||
Db 158 LQTVRRGLYLFSGPVGSG 176

RESULT 5
AAG92858
ID AAG92858 standard; protein; 229 AA.
XX
AC AAG92858;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum protein fragment SEQ ID NO: 6612.
XX
KW Corynebacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-00127688.
XX
PR 16-DEC-1999; 99JP-00377484.
PR 07-APR-2000; 2000JP-00159162.
PR 03-AUG-2000; 2000JP-00280988.
XX
PA (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayaashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI: 2001-376931/40.
DR N-PSDB; AAH68077.
XX
PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analyzing
PT expression profile or pattern of a gene and identifying homologous gene.
XX
PS Claim 17; SEQ ID NO 6612; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Corynebacterium glutamicum glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of corynebacterium bacterium, measuring expression amount and analysing
CC the expression profile or expression pattern of a gene derived from
CC Corynebacterium bacterium, and identifying a homologue of a gene derived from
CC corynebacterium bacterium. Corynebacterium bacteria are useful for producing amino
CC acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described in the
CC exemplification of the invention. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from the European Patent Office
XX
SQ Sequence 229 AA;
Query Match 43.8%; Score 45.5; DB 4; Length 229;
Best Local Similarity 47.6%; Pred. No. 49;
Matches 10; Conservative 4; Mismatches 4; Indels 3; Gaps 1;

QY 2 QRAVRPALSGF---DGRVSG 19
   ||| : ||| : |||
Db 168 QRATQAFTGYLTGDLGCG 188

RESULT 6
AAU43462
ID AAU43462 standard; protein; 67 AA.
XX
AC AAU43462;
XX
DT 27-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #4358.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US012865.
XX
PR 21-APR-2000; 2000US-0199047P.
PR 02-JUN-2000; 2000US-0208841P.
PR 07-JUL-2000; 2000US-0216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI: 2001-616774/71.
DR N-PSDB; AAS59520.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris.
XX
PS Example 1; SEQ ID NO 4657; 1069pp; English.
XX
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 67 AA;
Query Match 43.3%; Score 45; DB 4; Length 67;
Best Local Similarity 57.1%; Pred. No. 16;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 7 PALSGFGRVSGC 20
   | : | : | : |
Db 28 PCPAGSEGRVGLGC 41
```

RESULT 7
 ABM39981
 ID ABM39981 standard; protein; 67 AA.
 AC ABM39981;
 XX
 XX
 DT 20-OCT-2003 (first entry)
 XX
 DE Propionibacterium acnes predicted ORF-encoded polypeptide #4657.
 XX
 KW Acne vulgaris; antiseborrheic; dermatological; antibacterial;
 KW immunostimulant; immune response; vaccine.
 XX
 OS Propionibacterium acnes.
 XX
 XX
 PN WO2003033515-A1.
 XX
 PD 24-APR-2003.
 XX
 XX 11-OCT-2002; 2002WO-US032727.
 PF
 PF 15-OCT-2001; 2001US-00978825.
 PR
 XX (CORI-) CORIXA CORP.
 PA Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;
 PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;
 PI Earth B, Vallieue-Douglas J;
 XX
 XX WPI; 2003-381789/36.
 DR N-PSDB; ACF64449.
 XX
 XX New Propionibacterium acnes polypeptides and polynucleotides encoding the
 PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
 PT or for stimulating an immune response specific for a P. acnes protein.
 XX
 PS Example 1; SEQ ID NO 4657; 1481pp; English.
 XX
 XX The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
 CC encoding a Propionibacterium acnes protein. The invention also relates to
 CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
 CC immunogenic fragments of P. acnes polypeptides. The invention
 CC additionally encompasses expression vectors and host cells comprising a
 CC polynucleotide of the invention; antibodies against polypeptides of the
 CC invention; fusion proteins comprising a polypeptide of the invention; a
 CC method for stimulating an immune response specific for a P. acnes
 CC polypeptide and an isolated T cell population comprising T cells prepared
 CC via this method; a vaccine composition (comprising P. acnes polypeptides,
 CC polynucleotides, antibodies, fusion proteins, T cell populations, or
 CC antigen-presenting cells that express the polypeptide); a method and kit
 CC for detecting or determining the presence or absence of P. acnes in a
 CC patient; and a method for inhibiting the development of P. acnes in a
 CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
 CC proteins, T cell populations or antigen-presenting cells that express the
 CC polypeptides are useful for diagnosing, preventing or treating acne
 CC vulgaris, or for stimulating an immune response specific for a P. acnes
 CC protein. The polynucleotides can also be used as probes or primers for
 CC nucleic acid hybridization. The vaccine composition is useful for the
 CC stimulation of an immune response against P. acnes, or for treating acne,
 CC and the kit is useful for performing a diagnostic assay. The present
 CC sequence represents a polypeptide predicted to be encoded by an ORF (open
 CC reading frame) contained within the P. acnes polynucleotides of the
 CC invention. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 67 AA;

Query Match 43.3%; Score 45; DB 6; Length 67;
 Best Local Similarity 57.1%; Pred. No. 16;
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 7 PALSGFDGRVSGC 20
 Db 28 PCPAGSEGRVGLGC 41
 RESULT 8
 AAU63931
 ID AAU63931 standard; protein; 141 AA.
 XX
 AC AAU63931;
 XX
 DT 27-FEB-2002 (first entry)
 XX
 DE Propionibacterium acnes immunogenic protein #24827.
 XX
 KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KW dermatological; osteopathic; neuroprotectant.
 XX
 OS Propionibacterium acnes.
 XX
 PN WO200181581-A2.
 XX
 PD 01-NOV-2001.
 XX
 PF 20-APR-2001; 2001WO-US012865.
 XX
 PR 21-APR-2000; 2000US-0199047P.
 PR 02-JUN-2000; 2000US-0208841P.
 PR 07-JUL-2000; 2000US-0216747P.
 XX
 XX (CORI-) CORIXA CORP.
 PA
 PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
 XX
 XX WPI; 2001-616774/71.
 DR N-PSDB; AAS59637.
 XX
 XX Propionibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris.
 XX
 PS Example 1; SEQ ID NO 25126; 1069pp; English.
 XX
 XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
 CC polypeptides. The proteins and their associated DNA sequences are used in
 CC the treatment, prevention and diagnosis of medical conditions caused by
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
 CC P. acnes is also involved in infections of bone, joints and the central
 CC nervous system, however it is particularly involved in the inflammatory
 CC lesions associated with acne vulgaris. A method for detecting the
 CC presence or absence of P. acnes in a patient comprises contacting a
 CC sample with a binding agent that binds to the proteins of the invention
 CC and determining the amount of bound protein in the sample. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC specific for P. acnes proteins. These antibodies can be used to
 CC downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
 CC this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 141 AA;

Query Match 43.3%; Score 45; DB 4; Length 141;
 Best Local Similarity 55.6%; Pred. No. 36;
 Matches 10; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Matches 10; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3 RAVRPALSGDGRVSG 19
| | | | | | | | | |
Db 152 RRVTPLLRCRPGAGSG 168

RESULT 11
ABU48621
ID ABU48621 standard; protein; 287 AA.
XX
AC ABU48621;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #34148.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Treponema pallidum.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto RT, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR N-PSDB; ACA52491.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 76545; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,

CC K. pneumoniae or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 287 AA;
Query Match 43.3%; Score 45; DB 6; Length 287;
Best Local Similarity 36.8%; Pred. No. 75;
Matches 7; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 2 QRAVRPALSGDGRVSGC 20
::: : || || || ||
Db 186 KKSAGFIGGFAGSVGAGC 204

RESULT 12
AAU34790
ID AAU34790 standard; protein; 484 AA.
XX
AC AAU34790;
XX
DT 14-FEB-2002 (first entry)
XX
DE *E. coli* cellular proliferation protein #371.
XX
KW Antisense; prokaryotic cellular proliferation protein; antibiotic;
KW antibacterial; drug design.
XX
OS *Escherichia coli*.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US009180.
XX
PR 21-MAR-2000; 2000US-0191078P.
PR 23-MAY-2000; 2000US-0206848P.
PR 26-MAY-2000; 2000US-0207272P.
PR 23-OCT-2000; 2000US-0242578P.
PR 27-NOV-2000; 2000US-0253625P.
PR 22-DEC-2000; 2000US-0257931P.
PR 16-FEB-2001; 2001US-0269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR N-PSDB; AAS52649.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids.
XX
PS Example 3; SEQ ID NO 10383; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the genes,
CC their use in the discovery of novel antibiotics, the essential genes
CC themselves and the encoded proteins. The prokaryotes used are *Escherichia*
CC *coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*,
CC *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also
CC useful for the identification of potential new targets for antibiotic
CC development. The antisense nucleic acids can also be used to identify
CC proteins used in proliferation, to express these proteins, and to obtain
CC antibodies capable of binding to the expressed proteins. The proteins can
CC be used to screen compounds in rational drug discovery programmes. The
CC antisense nucleic acid sequence is also useful to screen for homologous
CC nucleic acids which are required for cell proliferation in a wide variety
CC of organisms. The present sequence represents an essential prokaryotic

CC cellular proliferation protein. Note: The sequence data for this patent
 CC did not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 484 AA;
 Query Match 43.3%; Score 45; DB 4; Length 484;
 Best Local Similarity 47.4%; Pred. No. 1.3e+02;
 Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;
 QY 2 QRAVRPALSGFDGRVSGC 20
 Db 87 QRVLRLPAILWMDGRCAQEC 105
 RESULT 13
 ABU28815
 ID ABU28815 standard; protein; 484 AA.
 AC ABU28815;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Protein encoded by Prokaryotic essential gene #14342.
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX
 OS Escherichia coli.
 XX
 PN WO200277183-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 21-MAR-2002; 2002WO-US009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX
 DR WPI; 2003-029926/02.
 DR N-PSDB; ACA32685.
 XX
 PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 PS Claim 25; SEQ ID NO 56739; 1766pp; English.
 XX
 CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a

CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 484 AA;
 Query Match 43.3%; Score 45; DB 6; Length 484;
 Best Local Similarity 47.4%; Pred. No. 1.3e+02;
 Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;
 QY 2 QRAVRPALSGFDGRVSGC 20
 Db 87 QRVLRLPAILWMDGRCAQEC 105
 RESULT 14
 ABP34437
 ID ABP34437 standard; protein; 73 AA.
 XX
 AC ABP34437;
 XX
 DT 08-JUL-2002 (first entry)
 XX
 DE Human ORF3410 protein, SEQ ID NO:6820.
 XX
 KW Human; ORF; open reading frame; ORFX; drug screening; diagnosis;
 KW disease monitoring; cytokine; cell proliferation; cell differentiation;
 KW immune modulation; hematopoiesis regulation; tissue growth;
 KW angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;
 KW thrombolytic; tumour inhibition; bodily characteristic; fertility;
 KW behaviour; cancer; proliferative disorder; neurological disorder;
 KW cardiovascular disease; immune system disorder; organ transplantation;
 KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
 KW hypothyroidism; cholesterol ester storage disease; infection; vulnery;
 KW vasotropic; antipsoriatic; antidiabetic; cycostatic; neurotropic;
 KW neuroprotective; antiatherosclerotic; anticoagulant; immunolytic;
 KW cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;
 KW dermatological; analgesic; virucide; antibacterial; fungicide.
 OS Homo sapiens.
 XX
 PN WO200190366-A2.
 XX
 PD 29-NOV-2001.
 XX
 PF 24-MAY-2001; 2001WO-US017076.
 XX
 PR 24-MAY-2000; 2000US-0206690P.
 XX
 PA (CURA-) CURAGEN CORP.
 PI Leach MD, Shinkets RA;
 XX
 DR WPI; 2002-106200/14.
 DR N-PSDB; ABN78463.
 XX
 PT Novel human polypeptides and polynucleotides useful for diagnosing,
 PT preventing and treating cardiovascular disease, neurodegenerative,
 PT hyperproliferative disorders and disorders related to organ
 PT transplantation.
 XX
 PS Claim 10; Page 1952; 2508pp; English.

CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 132 AA;

Query Match 42.3%; Score 44; DB 4; Length 132;
 Best Local Similarity 75.0%; Pred. No. 48;
 Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 7 PALSFGDGRVGS 18
 || : |||||
 DB 59 PASAGGDRVGS 70

RESULT 17
 ABG91056
 ID ABG91056 standard; protein; 943 AA.

XX AC ABG91056;

DT 29-NOV-2002 (first entry)

DE Neisseria gonorrhoeae lbpA protein.

XX Gram-negative bacterial bleb; PorB; outer membrane protein;
 KW Chlamydia trachomatis infection; Chlamydia pneumoniae infection;
 KW protective antigen; antibacterial; vaccine.

XX Neisseria gonorrhoeae.

PN WO200262380-A2.

XX PD 15-AUG-2002.

PF 08-FEB-2002; 2002WO-EP001356.

PR 08-FEB-2001; 2001GB-00003169.

PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

PI Berthet FJ, Lobet Y, Poolman J, Verlant VGCL;

XX WPI; 2002-657510/70.

DR N-PSDB; ABS67377.

XX Novel gram-negative bacterial bleb presenting on its surface PorB outer
 PT membrane protein from Chlamydia trachomatis or protective antigen from
 PT Chlamydia pneumoniae, useful for preventing Chlamydia infection.

PS Disclosure; Page 47-48; 75pp; English.

XX The present invention relates to a new gram-negative bacterial bleb
 CC presenting on its surface the PorB outer membrane protein from Chlamydia
 CC trachomatis, or a protective antigen from C. pneumoniae. The invention is
 CC useful for preventing C. trachomatis or C. pneumoniae infection in a
 CC host. The present amino acid sequence represents a Neisseria gonorrhoeae
 CC protein as described in the invention

XX Sequence 943 AA;

Query Match 42.3%; Score 44; DB 5; Length 943;

Best Local Similarity 60.0%; Pred. No. 3.8e+02;
 Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 5 VRPALSFGDGRVGS 19
 | : |||||
 DB 126 VQSLSGYGGRGSG 140

RESULT 18

AAW38756
 ID AAW38756 standard; protein; 1726 AA.

XX AC AAW38756;

XX 22-JUN-1998 (first entry)

XX Phosphatidyl inositol 3-kinase cdk-m.

XX Phosphatidyl inositol 3-kinase; signal transduction; cell cycle;
 KW antagonist; inflammatory joint disease; cell proliferation; cancer;
 KW psoriasis; restenosis; atherosclerosis; therapy; diagnosis; mouse; cpk-m.

XX Mus musculus.

XX Key Location/Qualifiers

FT Protein 69..1726

FT Peptide 391..404

FT Misc-difference 941

FT Domain 1175..1345

FT Domain 1590..1726

FT Domain 1590..1726

FT Domain 1590..1726

FT Domain 1590..1726

FT Domain 1590..1726

FT Domain 1590..1726

FT Domain 1590..1726

FT Domain 1590..1726

FT Domain 1590..1726

FT Domain 1590..1726

FT Domain 1590..1726

FT Domain 1590..1726

FT Domain 1590..1726

FT Domain 1590..1726

FT Domain 1590..1726

FT Domain 1590..1726

FT Domain 1590..1726

FT Domain 1590..1726

FT Domain 1590..1726

FT Domain 1590..1726

FT Domain 1590..1726

FT Domain 1590..1726

FT Domain 1590..1726

FT Domain 1590..1726

FT Domain 1590..1726

FT Domain 1590..1726

FT Domain 1590..1726

FT Domain 1590..1726

FT Domain 1590..1726

FT Domain 1590..1726

FT Domain 1590..1726

FT Domain 1590..1726

FT Domain 1590..1726

FT Domain 1590..1726

FT Domain 1590..1726

FT Domain 1590..1726

FT Domain 1590..1726

XX New isolated phosphatidyl inositol-3 kinase polypeptide - used to develop
 PT products for diagnosis and therapy, particularly for proliferative
 PT disorders, e.g. inflammatory joint diseases, or cancer.

XX Claim 5; Fig 10; 77pp; English.

XX This protein sequence comprises cpk-m, a murine polypeptide that belongs
 CC to a novel class of phosphatidyl inositol 3-kinases that contain a C2
 CC domain, are capable of phosphorylating a D3 hydroxyl of an inositol ring
 CC in phosphatidyl inositol and phosphatidyl inositol 4-phosphate, but not
 CC in phosphatidyl inositol 4,5-bisphosphate, and which are involved in cell
 CC signalling cascades that control e.g. cell cycle progression and
 CC intracellular protein sorting. The amino acid sequence was deduced from
 CC an isolated cDNA sequence (see AAT80199). It shows 34% identity and 48%
 CC similarity to Drosophila cpk (see AAW38757). Novel phosphatidyl inositol
 CC 3-kinases can be used to screen for agonists/antagonists of activity and
 CC in a claimed method of treating a disorder caused by dysregulation of a
 CC growth factor activation signalling cascade. Antagonists may reduce Ras
 CC activation allowing treatment of proliferative disorders such as
 CC atherosclerosis, inflammatory joint disease, psoriasis, restenosis
 CC following angioplasty, and cancer

XX

PS Claim 7; SEQ ID NO 30638; 1481pp; English.

XX The invention relates to an isolated polynucleotide (ACF64435-ACF64733) encoding a Propionibacterium acnes protein. The invention also relates to CC immunogenic fragments of P. acnes polypeptides (ABM35624-ABM64536) and to CC polynucleotides encoded by P. acnes polypeptides. The invention CC additionally encompasses expression vectors and host cells comprising a CC polynucleotide of the invention; antibodies against polypeptides of the CC invention; fusion proteins comprising a polypeptide of the invention; a CC method for stimulating an immune response specific for a P. acnes CC polypeptide and an isolated T cell population comprising T cells prepared CC via this method; a vaccine composition (comprising P. acnes polypeptides, CC polynucleotides, antibodies, fusion proteins, T cell populations, or CC antigen-presenting cells that express the polypeptide); a method and kit CC for detecting or determining the presence or absence of P. acnes in a CC patient; and a method for inhibiting the development of P. acnes in a CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion CC proteins, T cell populations or antigen-presenting cells that express the CC polypeptides are useful for diagnosing, preventing or treating acne CC vulgaris, or for stimulating an immune response specific for a P. acnes CC protein. The polynucleotides can also be used as probes or primers for CC nucleic acid hybridisation. The vaccine composition is useful for the CC stimulation of an immune response against P. acnes, or for treating acne, CC and the kit is useful for performing a diagnostic assay. The present CC sequence represents a specifically claimed P. acnes polypeptide which is CC thought to contain an immunogenic region. Note: The sequence data for CC this patent did not form part of the printed specification, but was CC obtained in electronic format directly from WIPO at CC ftp.wipo.int/pub/published_pct_sequences

XX Query Match 41.3%; Score 43; DB 6; Length 97;
Best Local Similarity 52.9%; Pred. No. 51;
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 LQRAVRPALSGFDGRVG 17
DQ 1 LSRLLHPVCSGRDRPG 17

RESULT 26
AAR98742
ID AAR98742 standard; protein; 228 AA.
XX AAR98742;
AC AAR98742;
XX 09-DEC-1996 (first entry)
DT Tazarotene induced gene 1 (TIG) encoded protein.
DE Tazarotene induced gene 1; TIG; retinoid; psoriasis; antipsoriatic;
KW fibroblast; keratinocyte; skin raft culture; retinoic acid receptor.
XX Homo sapiens.
OS Homo sapiens.
XX Key Location/Qualifiers
FH Domain 16..34
FT /label= Transmembrane_domain
FT /note= "putative transmembrane domain"
FT Modified-site 142..144
FT /label= Glycosylation
FT /note= "N-linked glycosylation site"
XX WO9623080-A1.
FN 01-AUG-1996.
PD 26-JAN-1996; 96WO-US001339.
PP 27-JAN-1995; 95US-00379280.
FR (ALLR) ALLERGAN.
XX PA

XX Nagpal S, Chandraratna RA;
PI WPI; 1996-362702/36.
XX N-PSDB; AAT38115.
PT Retinoid-inducible polynucleotide, designated Tazarotene Induced Gene
PT (TIG1) - having low basal expression, used in assay systems as an
PT indicator of retinoid action in psoriatic skin cultures.
XX Example 7; Page 40-41; 50pp; English.
XX A putative transmembrane protein (AAR98742) is encoded by novel human
CC TIG1 (Tazarotene induced gene 1) cDNA (AAT38115). TIG1 mRNA is strongly
CC induced from a low basal level upon treatment of skin raft cultures with
CC the retinoic acid receptor (RAR) beta/gamma selective antipsoriatic
CC synthetic retinoid AGN-190168 (ethyl 6-(2-(4,4) dimethyl-thiochroman-6-
CC yl) ethynyl-nicotinate). Inducible expression of TIG1 mRNA in psoriatic
CC skin raft cultures has also been demonstrated. The low basal expression
CC of TIG1 is partic. advantageous when used as an indicator of retinoid
CC action in psoriatic skin cultures
XX Sequence 228 AA;
SQ Query Match 41.3%; Score 43; DB 2; Length 228;
Best Local Similarity 61.1%; Pred. No. 1.2e+02;
Matches 11; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 LQRAVRPALSGFDGRVGS 18
DQ 59 LQAVRAALHFFNFRSGS 76

RESULT 27
ABR41629
ID ABR41629 standard; protein; 445 AA.
XX ABR41629;
AC ABR41629;
XX 02-JUN-2003 (first entry)
DT Human DITHP secreted/extracellular matrix protein.
DE Human; dithp; diagnostic and therapeutic polynucleotide; diagnosis;
KW cancer; cell proliferative disorder; autoimmune disorder;
KW inflammatory disorder; infection; hormonal disorder; metabolic disorder;
KW neurological disorder; gastrointestinal disorder; transport disorder;
KW connective tissue disorder; drug screening; proteome analysis;
KW gene therapy; antisense therapy; genotyping; transgenic animal; knock in;
KW disease model; toxicological testing; transcript imaging;
KW secreted protein; extracellular matrix.
XX Homo sapiens.
OS Homo sapiens.
XX WO200297031-A2.
PN 05-DEC-2002.
PD 27-MAR-2002; 2002WO-US010056.
PF 28-MAR-2001; 2001US-0279619P.
PR 29-MAR-2001; 2001US-0280067P.
PR 29-MAR-2001; 2001US-0280068P.
PR 16-MAY-2001; 2001US-0291280P.
PR 17-MAY-2001; 2001US-0291829P.
PR 17-MAY-2001; 2001US-0291849P.
PR 19-JUN-2001; 2001US-0299428P.
PR 20-JUN-2001; 2001US-0299776P.
PR 20-JUN-2001; 2001US-0300001P.
XX (INCY-) INCYTE GENOMICS INC.
PA Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
XX PI

PI Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Anshey SR;
 PI Daughtery SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;
 PI Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;
 PI Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;
 XX WPI: 2003-129518/12.
 DR N-PSDB; ACC46566.
 XX
 PT Novel human diagnostic and therapeutic polypeptide useful for identifying
 PT test compound which specifically binds to a polypeptide encoded by human
 PT diagnostic and therapeutic polynucleotide, and to induce antibodies.
 XX
 PS Claim 27; SEQ ID NO 1164; 591pp; English.
 XX
 CC The invention relates to novel human diagnostic and therapeutic
 CC polynucleotides designated dithp (ACC46080-ACC46749) and to their encoded
 CC proteins (DITHP; ABR41136-ABR41812). The invention also relates to
 CC polynucleotide sequences at least 90% identical to the dithp cDNA
 CC sequences of the invention; recombinant vectors, host cells and
 CC transgenic organisms comprising a dithp nucleic acid sequence; the
 CC recombinant production of DITHP proteins; antibodies specific for DITHP
 CC proteins; microarrays comprising dithp nucleic acid sequences; methods of
 CC detecting dithp nucleotide and protein sequences; methods of screening
 CC for compounds which specifically bind a DITHP protein; and methods of
 CC assessing the toxicity of test compounds using a dithp hybridisation
 CC probe. Dithp nucleic acid sequences and DITHP proteins may be used in the
 CC diagnosis of a wide variety of conditions including cancer and other cell
 CC proliferative disorders; autoimmune or inflammatory disorders; bacterial,
 CC viral, fungal or parasitic infections; hormonal disorders; metabolic
 CC disorders; neurological disorders; gastrointestinal disorders; transport
 CC disorders; and connective tissue disorders. They may also be used to
 CC screen for modulators of protein activity or gene expression. DITHP
 CC proteins can additionally be used in analysis of the proteome of a tissue
 CC or cell type and to induce antibodies. The dithp nucleic acids are
 CC additionally useful in somatic or germline gene therapy of the disorders
 CC mentioned above, as a source of antisense sequences, as a source of
 CC probes and primers, in genotyping and identification of individuals, in
 CC the generation of transgenic animal models of human disease or knock in
 CC humanised animals, in toxicological testing, and in transcript imaging.
 CC The present sequence represents a DITHP protein which is a secreted/
 CC extracellular matrix protein. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 445 AA;

Query Match 41.3%; Score 43; DB 6; Length 445;
 Best Local Similarity 58.3%; Pred. No. 2.5e+02;
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 5 VRPALSGDGRV 16
 Db 362 LRPALGLEGR 373

RESULT 28
 ABR41625
 ID ABR41625 standard; protein; 961 AA.
 AC ABR41625;
 AC ABR41625;
 XX
 DT 02-JUN-2003 (first entry)
 XX
 DE Human DITHP secreted/extracellular matrix protein.
 XX
 KW Human; dithp; diagnostic and therapeutic polynucleotide; diagnosis;
 KW cancer; cell proliferative disorder; autoimmune disorder;
 KW inflammatory disorder; infection; hormonal disorder; metabolic disorder;
 KW neurological disorder; gastrointestinal disorder; transport disorder;
 KW connective tissue disorder; drug screening; proteome analysis;
 KW gene therapy; antisense therapy; genotyping; transgenic animal; knock in;
 KW disease model; toxicological testing; transcript imaging;

secreted protein; extracellular matrix.

Homo sapiens.

WO200297031-A2.

05-DEC-2002.

27-MAR-2002; 2002WO-US010056.

28-MAR-2001; 2001US-0279619P.

29-MAR-2001; 2001US-0280067P.

29-MAR-2001; 2001US-0280068P.

16-MAY-2001; 2001US-0291280P.

17-MAY-2001; 2001US-0291829P.

19-JUN-2001; 2001US-0291849P.

20-JUN-2001; 2001US-0299428P.

20-JUN-2001; 2001US-0299776P.

(INCY-) INCVTE GENOMICS INC.

Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
 Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PE, Anshey SR;
 Daughtery SC, Dam TC, Liu TF, Nguyen DA, Kleefeld Y, Gerstin EH;
 Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;
 Flores V, Marwaha R, Lo A, Lan RY, Urashka ME;
 WPI: 2003-129518/12.
 N-PSDB; ACC46562.

Novel human diagnostic and therapeutic polypeptide useful for identifying
 test compound which specifically binds to a polypeptide encoded by human
 diagnostic and therapeutic polynucleotide, and to induce antibodies.

Claim 27; SEQ ID NO 1160; 591pp; English.

The invention relates to novel human diagnostic and therapeutic
 polynucleotides designated dithp (ACC46080-ACC46749) and to their encoded
 proteins (DITHP; ABR41136-ABR41812). The invention also relates to
 polynucleotide sequences at least 90% identical to the dithp cDNA
 sequences of the invention; recombinant vectors, host cells and
 transgenic organisms comprising a dithp nucleic acid sequence; the
 recombinant production of DITHP proteins; antibodies specific for DITHP
 proteins; microarrays comprising dithp nucleic acid sequences; methods of
 detecting dithp nucleotide and protein sequences; methods of screening
 for compounds which specifically bind a DITHP protein; and methods of
 assessing the toxicity of test compounds using a dithp hybridisation
 probe. Dithp nucleic acid sequences and DITHP proteins may be used in the
 diagnosis of a wide variety of conditions including cancer and other cell
 proliferative disorders; autoimmune or inflammatory disorders; bacterial,
 viral, fungal or parasitic infections; hormonal disorders; metabolic
 disorders; neurological disorders; gastrointestinal disorders; transport
 disorders; and connective tissue disorders. They may also be used to
 screen for modulators of protein activity or gene expression. DITHP
 proteins can additionally be used in analysis of the proteome of a tissue
 or cell type and to induce antibodies. The dithp nucleic acids are
 additionally useful in somatic or germline gene therapy of the disorders
 mentioned above, as a source of antisense sequences, as a source of
 probes and primers, in genotyping and identification of individuals, in
 the generation of transgenic animal models of human disease or knock in
 humanised animals, in toxicological testing, and in transcript imaging.
 The present sequence represents a DITHP protein which is a secreted/
 extracellular matrix protein. Note: The sequence data for this patent did
 not form part of the printed specification, but was obtained in
 electronic format directly from WIPO at
 ftp.wipo.int/pub/published_pct_sequences

Sequence 961 AA;

Query Match 41.3%; Score 43; DB 6; Length 961;

Best Local Similarity 58.3%; Pred. No. 5.6e+02;

Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

```

QY      5 VRPALSQFDGRV 16
Db      878 LRPALLEGEGRI 889

RESULT 29
AAG78887
ID AAG78887 standard; protein; 1121 AA.
XX
AC AAG78887;
XX
DT 14-DEC-2001 (first entry)
XX
DE Human fibrillin 3.
XX
KW Human; fibrillin; gene therapy; Marfan's syndrome.
XX
OS Homo sapiens.
XX
PN JP2001245664-A.
XX
PD 11-SEP-2001.
XX
PF 06-MAR-2000; 2000JP-00060009.
XX
PR 06-MAR-2000; 2000JP-00060009.
XX
PA (KAZU-) ZH KAZUSA DNA KENKYUSHO.
XX
DR WPI; 2001-610073/70.
DR N-PSDB; AAI69907.
XX
PT New human fibrillin 3 gene for treating and preventing diseases showing
PT symptoms similar to Marfan's syndrome.
XX
PS Claim 1; Page 11-14; 17pp; Japanese.
XX
CC The present sequence is the protein sequence for human fibrillin 3.
CC Fibrillin 3 and its coding sequence can be used for the treatment and the
CC prevention of diseases showing symptoms similar to Marfan's syndrome
XX
SQ Sequence 1121 AA;

Query Match 41.3%; Score 43; DB 4; Length 1121;
Best Local Similarity 58.3%; Pred. No. 6.6e+02;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY      5 VRPALSQFDGRV 16
Db      1038 LRPALLEGEGRI 1049

RESULT 30
ABB68757
ID ABB68757 standard; protein; 1169 AA.
XX
AC ABB68757;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 33063.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.

XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL12860.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
PS Disclosure; SEQ ID NO 33063; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLI16176-ABLI30511), expressed DNA
CC sequences (ABLI01840-ABLI16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1169 AA;

Query Match 41.3%; Score 43; DB 4; Length 1169;
Best Local Similarity 61.5%; Pred. No. 6.9e+02;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY      8 ALSGPDGRVSGGC 20
Db      961 ALCGFQGRLLAGC 973

Search completed: May 13, 2004, 06:52:31
Job time : 54 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2004, 06:49:48 ; Search time 14 Seconds
(without alignments)
73.751 Million cell updates/sec

Title: 09549186-7

Perfect score: 104

Sequence: 1 LQRAVRPALSGDFRVSVC 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Issued_Patents_AA.*

1: /cgn2_6/prodata/2/iaa/5A_COMB.pep.*

2: /cgn2_6/prodata/2/iaa/5B_COMB.pep.*

3: /cgn2_6/prodata/2/iaa/6A_COMB.pep.*

4: /cgn2_6/prodata/2/iaa/6B_COMB.pep.*

5: /cgn2_6/prodata/2/iaa/PCUTUS_COMB.pep.*

6: /cgn2_6/prodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	86	82.7	19	2	US-08-374-560-7
2	50	48.1	190	4	US-09-252-991A-19049
3	49	47.1	369	4	US-09-252-991A-27769
4	47.5	45.7	337	4	US-09-252-991A-19602
5	47	45.2	580	4	US-09-252-991A-20407
6	46.5	44.7	416	4	US-09-252-991A-30219
7	46	44.2	339	4	US-09-107-532A-3791
8	46	44.2	1073	4	US-09-252-991A-27341
9	45	43.3	172	4	US-09-800-729-188
10	44.5	42.8	341	4	US-09-252-991A-30646
11	44	42.3	263	4	US-09-489-039A-14057
12	44	42.3	435	4	US-09-489-039A-9982
13	44	42.3	607	3	US-08-537-361B-10
14	44	42.3	607	3	US-08-390-470A-4
15	44	42.3	607	3	US-08-817-707-10
16	44	42.3	780	4	US-09-252-991A-18846
17	44	42.3	941	3	US-09-074-658-75
18	44	42.3	944	2	US-08-867-941-23
19	44	42.3	944	2	US-08-867-941-24
20	44	42.3	944	3	US-09-074-658-23
21	44	42.3	944	3	US-09-074-658-24
22	44	42.3	1726	2	US-08-609-049A-30
23	44	42.3	1726	3	US-09-170-996-30
24	44	42.3	1778	4	US-09-252-991A-18159
25	44	42.3	1964	4	US-09-467-997-1
26	43.5	41.8	194	4	US-09-489-039A-11071
27	43.5	41.8	513	4	US-09-252-991A-21092

28	43	41.3	228	1	US-08-379-280-5	Sequence 5, Appli
29	42	40.4	99	4	US-08-311-731A-76	Sequence 76, Appl
30	42	40.4	115	4	US-08-311-731A-295	Sequence 295, App
31	42	40.4	155	4	US-09-328-352-7419	Sequence 7419, Ap
32	42	40.4	496	4	US-09-252-991A-20207	Sequence 20207, A
33	42	40.4	692	4	US-09-540-236-2750	Sequence 2750, Ap
34	42	40.4	818	4	US-09-252-991A-16691	Sequence 16691, A
35	42	40.4	898	2	US-08-867-941-11	Sequence 11, Appl
36	42	40.4	2439	3	US-09-074-658-11	Sequence 11, Appl
37	41.5	39.9	101	4	US-08-469-260A-60	Sequence 60, Appl
38	41.5	39.9	101	4	US-08-488-446-60	Sequence 60, Appl
39	41.5	39.9	101	4	US-08-467-344A-60	Sequence 60, Appl
40	41.5	39.9	660	4	US-09-252-991A-32206	Sequence 32206, A
41	41.5	39.9	984	4	US-09-287-354-2	Sequence 2, Appli
42	41.5	39.9	1189	4	US-09-287-354-3	Sequence 3, Appli
43	41.5	39.9	1189	4	US-09-287-354-4	Sequence 4, Appli
44	41	39.4	134	5	PCN-US95-06266-146	Sequence 146, App
45	41	39.4	175	4	US-09-252-991A-23774	Sequence 23774, A
46	41	39.4	284	4	US-09-252-991A-25592	Sequence 25592, A
47	41	39.4	297	4	US-09-252-991A-26587	Sequence 26587, A
48	41	39.4	415	4	US-09-252-991A-26760	Sequence 26760, A
49	41	39.4	492	4	US-09-252-991A-28339	Sequence 28339, A
50	41	39.4	894	2	US-08-867-941-15	Sequence 15, Appl
51	41	39.4	2432	3	US-09-074-658-15	Sequence 15, Appl
52	40.5	38.9	436	4	US-09-252-991A-20256	Sequence 20256, A
53	40.5	38.9	509	4	US-09-043-302-5	Sequence 5, Appli
54	40.5	38.9	509	4	US-09-043-302-11	Sequence 11, Appl
55	40	38.5	134	4	US-09-252-991A-23103	Sequence 23103, A
56	40	38.5	157	4	US-09-252-991A-20241	Sequence 20241, A
57	40	38.5	259	4	US-09-489-039A-9829	Sequence 9829, Ap
58	40	38.5	432	4	US-09-252-991A-31177	Sequence 31177, A
59	40	38.5	448	4	US-09-215-418-4	Sequence 4, Appli
60	40	38.5	483	4	US-09-252-991A-20504	Sequence 20504, A
61	40	38.5	491	4	US-09-252-991A-24224	Sequence 24224, A
62	40	38.5	742	4	US-09-252-991A-24289	Sequence 24289, A
63	40	38.5	937	4	US-09-252-991A-30507	Sequence 30507, A
64	39.5	38.0	309	4	US-09-252-991A-17744	Sequence 17744, A
65	39.5	38.0	443	4	US-09-252-991A-18775	Sequence 18775, A
66	39.5	38.0	450	4	US-09-252-991A-32441	Sequence 32441, A
67	39.5	38.0	489	4	US-09-252-991A-27833	Sequence 27833, A
68	39.5	38.0	538	4	US-09-252-991A-21622	Sequence 21622, A
69	39.5	38.0	580	4	US-09-252-991A-30180	Sequence 30180, A
70	39.5	38.0	802	4	US-09-252-991A-25050	Sequence 25050, A
71	39	37.5	112	4	US-09-087-031E-23	Sequence 23, Appl
72	39	37.5	119	4	US-09-252-991A-20541	Sequence 20541, A
73	39	37.5	134	5	PCN-US95-06266-144	Sequence 144, App
74	39	37.5	155	4	US-09-252-991A-18135	Sequence 18135, A
75	39	37.5	159	4	US-09-252-991A-30157	Sequence 30157, A
76	39	37.5	166	4	US-09-621-976-4135	Sequence 4135, Ap
77	39	37.5	183	4	US-09-252-991A-28717	Sequence 28717, A
78	39	37.5	212	4	US-09-252-991A-27887	Sequence 27887, A
79	39	37.5	259	4	US-09-252-991A-28573	Sequence 28573, A
80	39	37.5	292	4	US-09-818-780-10	Sequence 10, Appl
81	39	37.5	450	4	US-09-489-039A-13258	Sequence 13258, A
82	39	37.5	467	4	US-09-252-991A-20689	Sequence 20689, A
83	39	37.5	470	4	US-09-252-991A-23310	Sequence 23310, A
84	39	37.5	470	4	US-09-328-352-6912	Sequence 6912, Ap
85	39	37.5	500	4	US-09-252-991A-21214	Sequence 21214, A
86	39	37.5	641	4	US-09-252-991A-26329	Sequence 26329, A
87	38.5	37.0	51	4	US-09-621-976-5922	Sequence 5922, Ap
88	38.5	37.0	234	4	US-09-489-039A-13353	Sequence 13353, A
89	38.5	37.0	382	4	US-09-252-991A-25262	Sequence 25262, A
90	38.5	37.0	425	1	US-08-615-170-17	Sequence 17, Appl
91	38.5	37.0	432	1	US-08-615-170-3	Sequence 3, Appli
92	38.5	37.0	445	1	US-08-615-170-5	Sequence 5, Appli
93	38.5	37.0	493	4	US-09-252-991A-27576	Sequence 27576, A
94	38.5	37.0	770	4	US-09-252-991A-17062	Sequence 17062, A
95	38.5	37.0	832	4	US-09-252-991A-19252	Sequence 19252, A
96	38.5	37.0	947	4	US-09-252-991A-25382	Sequence 25382, A
97	38.5	37.0	1495	4	US-09-543-681A-5986	Sequence 5986, Ap
98	38	36.5	14	1	US-07-854-603-8	Sequence 8, Appli
99	38	36.5	150	4	US-09-489-039A-10878	Sequence 10878, A
100	38	36.5	194	4	US-09-252-991A-28941	Sequence 28941, A

ALIGNMENTS

RESULT 1

US-08-374-560-7
; Sequence 7, Application US/08374560
; Patent No. 5882645
; GENERAL INFORMATION:
; APPLICANT: TOTH, Istvan
; APPLICANT: GIBBONS, William Anthony
; TITLE OF INVENTION: PEPTIDE COMPOUNDS
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen
; STREET: 1180 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10036-8403
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/374,560
; FILING DATE: 13-MAR-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9215780.9
; FILING DATE: 24-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Meilman, Edward A.
; REGISTRATION NUMBER: 24,735
; REFERENCE/DOCKET NUMBER: P/365-302
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 382-0700
; TELEFAX: (212) 382-0888
; TELEX: 236925
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
US-08-374-560-7

Query Match 82.7%; Score 86; DB 2; Length 19;
Best Local Similarity 100.0%; Pred. No. 8.2e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RAVRPALSGDGRVGS 19
Db 3 RAVRPALSGDGRVGS 19

RESULT 2

US-09-252-991A-19049
; Sequence 19049, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19049
; LENGTH: 190
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19049

Query Match 48.1%; Score 50; DB 4; Length 190;
Best Local Similarity 56.2%; Pred. No. 1.1;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 ORAVRPALSGDGRVG 17
Db 100 RCVRPAAAGGPGRAG 115

RESULT 3

US-09-252-991A-27769
; Sequence 27769, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27769
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27769

Query Match 47.1%; Score 49; DB 4; Length 369;
Best Local Similarity 60.0%; Pred. No. 3.3;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 6 RPALSGDGRVGS 20
Db 4 RPDGLGCGRYGAGC 18

RESULT 4

US-09-252-991A-19602
; Sequence 19602, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19602
; LENGTH: 337
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19602

Query Match 45.7%; Score 47.5; DB 4; Length 337;
Best Local Similarity 61.9%; Pred. No. 5.3;


```
Matches 13; Conservative 1; Mismatches 4; Indels 3; Gaps 1;
QY 2 QRAVRPAL---SGFDGRVGS 19
:|||||:|||||
Db 294 RRADRPALAPGSGFAGRVVG 314

RESULT 5
US-09-252-991A-20407
; Sequence 20407, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 20407
; LENGTH: 580
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20407

Query Match 45.2%; Score 47; DB 4; Length 580;
Best Local Similarity 55.8%; Pred. No. 12;
Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
QY 2 QRAVRPALSGFDGRVGS 19
:|||||:|||||
Db 87 EQVRPALPGGGAVGHG 104

RESULT 6
US-09-252-991A-30219
; Sequence 30219, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30219
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30219

Query Match 44.7%; Score 46.5; DB 4; Length 416;
Best Local Similarity 68.8%; Pred. No. 9.9;
Matches 11; Conservative 1; Mismatches 3; Indels 1; Gaps 1;
QY 3 RAVRPALSGFDGRVG 17
:|||||:|||||
Db 203 RAVRPAAAGHDGRTG 218

RESULT 7
US-09-107-532A-3791
; Sequence 3791, Application US/09107532A
; Patent No. 6583275
```

```
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneka
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3791:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 339 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...339
; SEQUENCE DESCRIPTION: SEQ ID NO: 3791:
US-09-107-532A-3791

Query Match 44.2%; Score 46; DB 4; Length 339;
Best Local Similarity 57.9%; Pred. No. 9.5;
Matches 11; Conservative 1; Mismatches 7; Indels 0; Gaps 0;
QY 1 LQRAVRPALSGFDGRVGS 19
:|||||:|||||
Db 158 LQRTVRGLYLFSGPVGSG 176

RESULT 8
US-09-252-991A-27341
; Sequence 27341, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27341
```

; LENGTH: 1073
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (803)
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
US-09-252-991A-27341

Query Match 44.2%; Score 46; DB 4; Length 1073;
Best Local Similarity 69.2%; Pred. No. 35;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 6 RPALSGDGRVGS 18
DB 777 RPLAGPDGRAGS 789

RESULT 9
US-09-800-729-188
; Sequence 188, Application US/09800729
; Patent No. 6605592
; GENERAL INFORMATION:
; APPLICANT: Ni et al.
; TITLE OF INVENTION: 32 Human secreted proteins
; FILE REFERENCE: P204491
; CURRENT APPLICATION NUMBER: US/09/800,729
; CURRENT FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: PCT/US00/26013
; PRIOR FILING DATE: 2000-09-22
; PRIOR APPLICATION NUMBER: 60/155,709
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 217
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 188
; LENGTH: 172
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-800-729-188

Query Match 43.3%; Score 45; DB 4; Length 172;
Best Local Similarity 58.8%; Pred. No. 6.5;
Matches 10; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3 RAVRPALSGDGRVGS 19
DB 152 RRVTPLLGRGRAGSG 168

RESULT 10
US-09-252-991A-30646
; Sequence 30646, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30646
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30646
Query Match 42.8%; Score 44.5; DB 4; Length 341;
Best Local Similarity 42.3%; Pred. No. 17;

Matches 11; Conservative 2; Mismatches 4; Indels 9; Gaps 1;
QY 4 AVRPALSGDGR-----VGS GC 20
DB 1 AVROAQRGRDRRRSCLWGQYIGNGC 26

RESULT 11
US-09-489-039A-14057
; Sequence 14057, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 14057
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-14057

Query Match 42.3%; Score 44; DB 4; Length 263;
Best Local Similarity 64.3%; Pred. No. 15;
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 6 RPALSGDGRVGS 19
DB 59 RPAARGFSGSGSG 72

RESULT 12
US-09-489-039A-9982
; Sequence 9982, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 9982
; LENGTH: 435
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-9982

Query Match 42.3%; Score 44; DB 4; Length 435;
Best Local Similarity 50.0%; Pred. No. 27;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 QRAVRPALSGDGRVGS 19
DB 281 QRVGHHLLGGFKRGLNG 298

RESULT 13
US-08-537-361E-10
; Sequence 10, Application US/08537361E
; Patent No. 6121037
; GENERAL INFORMATION:
; APPLICANT: Stojiljkovic, Igor
; APPLICANT: SO, Magdalene
; APPLICANT: Hwa, Vivian

```

; APPLICANT: Heffron, Fred
; APPLICANT: Nassif, Xavier
; TITLE OF INVENTION: No. 6121037el Bacterial Hemoglobin Receptor
; TITLE OF INVENTION: Genes and Uses
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive, 32nd Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/537,361E
; FILING DATE: 02-OCT-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6121037nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 94,784-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0002
; TELEFAX: 312-913-0001
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 607 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-537-361E-10

Query Match 42.3%; Score 44; DB 3; Length 607;
Best Local Similarity 60.0%; Pred. No. 39;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 5 VRPALSQFDGRVGS 19
Db 126 VQSLSGYGGRGSG 140

RESULT 14
US-08-990-470A-4
; Sequence 4, Application US/08990470A
; Patent No. 6123942
; GENERAL INFORMATION:
; APPLICANT: Stojiljkovic, Igor
; APPLICANT: So, Magdalene
; APPLICANT: Hwa, Vivian
; APPLICANT: Heffron, Fred
; APPLICANT: Nassif, Xavier
; TITLE OF INVENTION: No. 6123942el Bacterial Hemoglobin Receptor
; TITLE OF INVENTION: Genes and Uses
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive, 32nd Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/990,470A
; FILING DATE: 15-DEC-1997

```

```

; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6123942nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 94,784-K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 607 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-990-470A-4

Query Match 42.3%; Score 44; DB 3; Length 607;
Best Local Similarity 60.0%; Pred. No. 39;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 5 VRPALSQFDGRVGS 19
Db 126 VQSLSGYGGRGSG 140

RESULT 15
US-08-817-707-10
; Sequence 10, Application US/08817707
; Patent No. 6277382
; GENERAL INFORMATION:
; APPLICANT: Stojiljkovic, Igor
; APPLICANT: So, Magdalene
; APPLICANT: Hwa, Vivian
; APPLICANT: Heffron, Fred
; APPLICANT: Nassif, Xavier
; TITLE OF INVENTION: No. 6277382el Bacterial Hemoglobin Receptor
; TITLE OF INVENTION: Genes and Uses
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive, 32nd Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,707
; FILING DATE: 19-AUG-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6277382nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 94,784-J
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 607 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-817-707-10

Query Match 42.3%; Score 44; DB 3; Length 607;
Best Local Similarity 60.0%; Pred. No. 39;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

```

QY 5 VRPALSFGDGRVSG 19
; : : : :
Db 126 VQGSLSGYGGRGSG 140
; : : : :
US-09-074-658-75

RESULT 16
US-09-252-991A-18846
; Sequence 18846, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18846
; LENGTH: 780
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18846

Query Match 42.3%; Score 44; DB 4; Length 780;
Best Local Similarity 64.3%; Pred. No. 52;
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 7 PALSFGDGRVSGC 20
; : : : :
Db 157 PARPGAGRRHSGC 170
; : : : :
US-09-074-658-75

RESULT 17
US-09-074-658-75
; Sequence 75, Application US/09074658
; Patent No. 6184371
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M
; APPLICANT: Run-Pan Du
; APPLICANT: Quijun Wang
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA
; NUMBER OF SEQUENCES: 78
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/074,658
; FILING DATE: 08-MAY-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-795
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 944 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-074-658-75

Query Match 42.3%; Score 44; DB 4; Length 944;
Best Local Similarity 60.0%; Pred. No. 64;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 5 VRPALSFGDGRVSG 19
; : : : :
Db 127 VQGSLSGYGGRGSG 141
; : : : :
US-08-867-941-24

RESULT 19
US-08-867-941-24
; Sequence 24, Application US/08867941
; Patent No. 5977337
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M

; LENGTH: 941 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-074-658-75

Query Match 42.3%; Score 44; DB 3; Length 941;
Best Local Similarity 60.0%; Pred. No. 64;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 5 VRPALSFGDGRVSG 19
; : : : :
Db 122 VQGSLSGYGGRGSG 136
; : : : :
US-08-867-941-23

RESULT 18
US-08-867-941-23
; Sequence 23, Application US/08867941
; Patent No. 5977337
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M
; APPLICANT: Du, Run-Pan
; APPLICANT: Wang, Quijun
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/867,941
; FILING DATE: 03-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-681 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 944 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-867-941-23

Query Match 42.3%; Score 44; DB 2; Length 944;
Best Local Similarity 60.0%; Pred. No. 64;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 5 VRPALSFGDGRVSG 19
; : : : :
Db 127 VQGSLSGYGGRGSG 141
; : : : :
US-08-867-941-24

RESULT 19
US-08-867-941-24
; Sequence 24, Application US/08867941
; Patent No. 5977337
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M

APPLICANT: Du, Run-Pan
APPLICANT: Wang, Quijun
APPLICANT: Yang, Yan-Ping
APPLICANT: Klein, Michel H
TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA
NUMBER OF SEQUENCES: 67
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/867,941
FILING DATE: 03-JUN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-681 MIS:jb
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 944 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-867-941-24

Query Match 42.3%; Score 44; DB 2; Length 944;
Best Local Similarity 60.0%; Pred. No. 64;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 5 VRPALSFGDGRVGS 19
|: |||: |||
DB 127 VQSLSGYGRGSG 141

RESULT 20
US-09-074-658-23
Sequence 23, Application US/09074658
Patent No. 6184371
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena M
APPLICANT: Run-Pan Du
APPLICANT: Quijun Wang
APPLICANT: Yang, Yan-Ping
APPLICANT: Klein, Michel H
TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA
NUMBER OF SEQUENCES: 78
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/074,658
FILING DATE: 08-MAY-1998

Query Match 42.3%; Score 44; DB 2; Length 944;
Best Local Similarity 60.0%; Pred. No. 64;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 5 VRPALSFGDGRVGS 19
|: |||: |||
DB 127 VQSLSGYGRGSG 141

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-795
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 944 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-074-658-23

Query Match 42.3%; Score 44; DB 3; Length 944;
Best Local Similarity 60.0%; Pred. No. 64;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 5 VRPALSFGDGRVGS 19
|: |||: |||
DB 127 VQSLSGYGRGSG 141

RESULT 21
US-09-074-658-24
Sequence 24, Application US/09074658
Patent No. 6184371
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena M
APPLICANT: Run-Pan Du
APPLICANT: Quijun Wang
APPLICANT: Yang, Yan-Ping
APPLICANT: Klein, Michel H
TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA
NUMBER OF SEQUENCES: 78
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/074,658
FILING DATE: 08-MAY-1998

Query Match 42.3%; Score 44; DB 3; Length 944;
Best Local Similarity 60.0%; Pred. No. 64;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 5 VRPALSFGDGRVGS 19
|: |||: |||
DB 127 VQSLSGYGRGSG 141

```

; : : : : :
Db 127 VQGSLSGVGGRGSG 141

RESULT 22
US-08-609-049A-30
; Sequence 30, Application US/08609049A
; Patent No. 5948664
; GENERAL INFORMATION:
; APPLICANT: Williams, Lewis T.
; APPLICANT: Molz, Lisa
; APPLICANT: Chen, Yen-Wen
; TITLE OF INVENTION: No. 5948664e1 PI 3-Kinase Polypeptides
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/609,049A
; FILING DATE: 29-FEB-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Karen B.
; REGISTRATION NUMBER: 29,684
; REFERENCE/DOCKET NUMBER: 2307K-063700US
; TELEPHONE: 415-326-2422
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1726 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-609-049A-30

Query Match 42.3%; Score 44; DB 2; Length 1726;
Best Local Similarity 44.0%; Pred. No. 1.3e+02;
Matches 11; Conservative 1; Mismatches 7; Indels 6; Gaps 1;

QY 2 QRAVRPALSG-----FDGRVSGC 20
Db 14 QRPAEPAEAGEKHGSDLGREGSGC 38

RESULT 24
US-09-252-991A-18159
; Sequence 18159, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18159
; LENGTH: 1778
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18159

Query Match 42.3%; Score 44; DB 4; Length 1778;
Best Local Similarity 60.0%; Pred. No. 1.3e+02;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 AVRPAALSGFDGRVGS 18
Db 461 ARQPASGFLGRIGN 475

RESULT 25
US-09-467-997-1
; Sequence 1, Application US/09467997
; Patent No. 6379925
; GENERAL INFORMATION:
; APPLICANT: Kitajewski, Jan
; APPLICANT: Uyttendaele, Hendrik
; TITLE OF INVENTION: ANGIOGENIC MODULATION BY NOTCH SIGNAL TRANSDUCTION
```

FILE REFERENCE: 53863-A-PCT-US
CURRENT APPLICATION NUMBER: US/09/467,997
CURRENT FILING DATE: 1999-12-20
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 1964
TYPE: PRT
ORGANISM: mouse
US-09-467-997-1

Query Match 42.3%; Score 44; DB 4; Length 1964;
Best Local Similarity 52.9%; Pred. No. 1.5e+02;
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 3 RAVRPALSGDGRVSGG 19
| | | | | | | | | |
Db 1165 RCORPGASCGRGGDG 1181

RESULT 26
US-09-489-039A-11071
Sequence 11071, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
FILE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 2709.2004001
CURRENT APPLICATION NUMBER: US/09/489,039A
CURRENT FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 11071
LENGTH: 194
TYPE: PRT
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-11071

Query Match 41.8%; Score 43.5; DB 4; Length 194;
Best Local Similarity 31.4%; Pred. No. 13;
Matches 11; Conservative 2; Mismatches 7; Indels 15; Gaps 1;

QY 1 LQAVRPALSGDGRV-----GSGC 20
| | | | | | | | | |
Db 118 LQSQINPQLAGHGRVSLMEITDDGLAILQFGGCG 152

RESULT 27
US-09-252-991A-21092
Sequence 21092, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 21092
LENGTH: 513
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21092

Query Match 41.8%; Score 43.5; DB 4; Length 513;
Best Local Similarity 45.8%; Pred. No. 40;

Matches 11; Conservative 1; Mismatches 7; Indels 5; Gaps 1;

QY 2 QRAVRPALSG-----FDGRVSGGC 20
| | | | | | | | | |
Db 229 QHAVRPDARGHLVDVSHDGRHAAGC 252

RESULT 28
US-08-379-280-5
Sequence 5, Application US/08379280
Patent No. 5650279
GENERAL INFORMATION:
APPLICANT: Negpal, Sunil
APPLICANT: Chandraratna, Roshantha A. S.
TITLE OF INVENTION: GENE SEQUENCE INDUCED IN SKIN BY RETINOIDS
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: CA
COUNTRY: U.S.A.
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/379,280
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Altman, Daniel E
REGISTRATION NUMBER: 34,115
REFERENCE/DOCKET NUMBER: ALRGN.057A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 714-760-0404
TELEFAX: 714-760-9502
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 228 amino acids
TYPE: amino acids
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
US-08-379-280-5

Query Match 41.3%; Score 43; DB 1; Length 228;
Best Local Similarity 61.1%; Pred. No. 19;
Matches 11; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 LQAVRPALSGDGRVGS 18
| | | | | | | | | |
Db 59 LQAVRAALHFFNFRSGS 76

RESULT 29
US-08-311-731A-76
Sequence 76, Application US/08311731A
Patent No. 6583266
GENERAL INFORMATION:
APPLICANT: SMITH, DOUGLAS
APPLICANT: MAO, JEN-I
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES

;/ TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
;/ TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
;/ NUMBER OF SEQUENCES: 411
;/ CORRESPONDENCE ADDRESS:
;/ ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
;/ STREET: 600 ATLANTIC AVENUE
;/ CITY: BOSTON
;/ STATE: MASSACHUSETTS
;/ COUNTRY: USA
;/ ZIP: 02210
;/ COMPUTER READABLE FORM:
;/ MEDIUM TYPE: Floppy disk
;/ COMPUTER: IBM PC compatible
;/ OPERATING SYSTEM: PC-DOS/MS-DOS
;/ SOFTWARE: Patent In Release #1.0, Version #1.25
;/ CURRENT APPLICATION DATA:
;/ APPLICATION NUMBER: US/08/311,731A
;/ FILING DATE:
;/ CLASSIFICATION: 530
;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: GATES, EDWARD R.
;/ REGISTRATION NUMBER: 31,616
;/ REFERENCE/DOCKET NUMBER: C0044/7125
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: 617/720-3500
;/ TELEFAX: 617/720-2441
;/ INFORMATION FOR SEQ ID NO: 76:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 99 amino acids
;/ TYPE: amino acid
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: protein
;/ HYPOTHETICAL: YES
;/ ORIGINAL SOURCE:
;/ ORGANISM: MYCOBACTERIUM LEPRAE
;/ US-08-311-731A-76

Query Match 40.4%; Score 42; DB 4; Length 99;
Best Local Similarity 50.0%; Pred. No. 11;
Matches 10; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 LQAVRPALSGFDGRVSGC 20
Db 41 LSYAVSSGPGFKGRGRC 60

RESULT 30

US-08-311-731A-295
;/ Sequence 295, Application US/08311731A
;/ Patent No. 6583266
;/ GENERAL INFORMATION:
;/ APPLICANT: SMITH, DOUGLAS
;/ APPLICANT: MAO, JEN-I
;/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
;/ TITLE OF INVENTION: RELATING TO MYCOBACTERIUM TUBERCULOSIS AND LAPRAE FOR
;/ TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
;/ NUMBER OF SEQUENCES: 411
;/ CORRESPONDENCE ADDRESS:
;/ ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
;/ STREET: 600 ATLANTIC AVENUE
;/ CITY: BOSTON
;/ STATE: MASSACHUSETTS
;/ COUNTRY: USA
;/ ZIP: 02210
;/ COMPUTER READABLE FORM:
;/ MEDIUM TYPE: Floppy disk
;/ COMPUTER: IBM PC compatible
;/ OPERATING SYSTEM: PC-DOS/MS-DOS
;/ SOFTWARE: Patent In Release #1.0, Version #1.25
;/ CURRENT APPLICATION DATA:
;/ APPLICATION NUMBER: US/08/311,731A
;/ FILING DATE:
;/ CLASSIFICATION: 530

;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: GATES, EDWARD R.
;/ REGISTRATION NUMBER: 31,616
;/ REFERENCE/DOCKET NUMBER: C0044/7125
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: 617/720-3500
;/ TELEFAX: 617/720-2441
;/ INFORMATION FOR SEQ ID NO: 295:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 115 amino acids
;/ TYPE: amino acid
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: protein
;/ HYPOTHETICAL: YES
;/ ORIGINAL SOURCE:
;/ ORGANISM: Mycobacterium leprae
;/ US-08-311-731A-295

Query Match 40.4%; Score 42; DB 4; Length 115;
Best Local Similarity 50.0%; Pred. No. 13;
Matches 10; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 LQAVRPALSGFDGRVSGC 20
Db 57 LSYAVSSGPGFKGRGRC 76

Search completed: May 13, 2004, 06:56:38
Job time : 16 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2004, 06:55:22 ; Search time 214 Seconds
(without alignments)
26.006 Million cell updates/sec

Title: 09549186-7
Perfect score: 104
Sequence: 1 LQRAVRPALSGDFGRVSGC 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1145568 seqs, 278261457 residues

Total number of hits satisfying chosen parameters: 1145568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Published Applications AA.*
1: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/prodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/prodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/prodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/prodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/prodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/prodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/prodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	46	44.2	404	12	US-10-425-114-71851
2	45.5	43.8	229	9	US-09-738-626-6612
3	45	43.3	148	12	US-10-424-599-263786
4	45	43.3	162	12	US-10-425-114-50038
5	45	43.3	172	9	US-09-800-729-188
6	45	43.3	233	12	US-10-425-114-72819
7	45	43.3	247	12	US-10-425-114-38641
8	45	43.3	287	12	US-10-282-122A-76545
9	45	43.3	299	15	US-10-369-493-11875
10	45	43.3	484	9	US-09-815-242-10383
11	45	43.3	484	12	US-10-282-122A-56739
12	45	43.3	484	15	US-10-369-493-23598
13	44.5	42.8	73	11	US-09-864-408A-6820
14	44.5	42.8	99	12	US-10-424-599-150057
15	44.5	42.8	301	15	US-10-369-493-1228
					Sequence 71851, A
					Sequence 6612, Ap
					Sequence 263786,
					Sequence 50038, A
					Sequence 188, App
					Sequence 72819, A
					Sequence 38641, A
					Sequence 76545, A
					Sequence 11875, A
					Sequence 10383, A
					Sequence 56739, A
					Sequence 23598, A
					Sequence 6820, Ap
					Sequence 150057,
					Sequence 1228, Ap

16	44	42.3	133	12	US-10-424-599-177904	Sequence 177904,
17	44	42.3	184	15	US-10-369-493-3741	Sequence 3741, Ap
18	44	42.3	364	12	US-10-424-599-198861	Sequence 198861,
19	43.5	41.8	111	12	US-10-424-599-241830	Sequence 241830, A
20	43	41.3	179	12	US-10-425-114-46577	Sequence 46577, A
21	43	41.3	298	14	US-10-156-761-10036	Sequence 10036, A
22	43	41.3	299	12	US-10-425-114-44283	Sequence 44283, A
23	43	41.3	2695	15	US-10-015-115-14	Sequence 14, Appl
24	43	41.3	2757	15	US-10-015-115-16	Sequence 16, Appl
25	43	41.3	2844	15	US-10-015-115-8	Sequence 8, Appl
26	43	41.3	2845	15	US-10-015-115-12	Sequence 12, Appl
27	43	41.3	2877	15	US-10-015-115-10	Sequence 10, Appl
28	43	41.3	2995	15	US-10-015-115-6	Sequence 6, Appl
29	42.5	40.9	87	12	US-10-424-599-275944	Sequence 275944,
30	42.5	40.9	543	15	US-10-369-493-19465	Sequence 19465, A
31	42	40.4	243	12	US-10-412-699B-1447	Sequence 1447, Ap
32	42	40.4	243	15	US-10-310-154-642	Sequence 642, App
33	42	40.4	243	15	US-10-374-780A-1302	Sequence 1302, Ap
34	42	40.4	263	12	US-10-282-122A-60790	Sequence 60790, A
35	42	40.4	273	12	US-10-282-122A-49486	Sequence 49486, A
36	42	40.4	551	12	US-10-282-122A-49838	Sequence 49838, A
37	42	40.4	608	16	US-10-389-566-752	Sequence 752, App
38	42	40.4	616	15	US-10-369-493-17046	Sequence 17046, A
39	42	40.4	1399	14	US-10-156-761-14753	Sequence 14753, A
40	41.5	39.9	101	8	US-08-424-508B-60	Sequence 60, Appl
41	41.5	39.9	343	10	US-09-866-050A-656	Sequence 656, App
42	41.5	39.9	396	12	US-10-282-122A-76600	Sequence 76600, A
43	41.5	39.9	984	14	US-10-024-368-2	Sequence 2, Appl
44	41.5	39.9	986	12	US-10-282-122A-42808	Sequence 42808, A
45	41.5	39.9	1189	14	US-10-024-368-3	Sequence 3, Appl
46	41.5	39.9	1189	14	US-10-024-368-4	Sequence 4, Appl
47	41.5	39.9	1189	14	US-10-122-013-17	Sequence 17, Appl
48	41	39.4	68	12	US-10-424-599-227312	Sequence 227312,
49	41	39.4	135	12	US-10-425-114-68463	Sequence 68463, A
50	41	39.4	208	14	US-10-017-161-1888	Sequence 1888, Ap
51	41	39.4	208	15	US-10-292-798-1544	Sequence 1544, Ap
52	41	39.4	273	15	US-10-369-493-1170	Sequence 1170, Ap
53	41	39.4	291	12	US-10-425-114-71228	Sequence 71228, A
54	41	39.4	388	15	US-10-369-493-18132	Sequence 18132, A
55	41	39.4	473	15	US-10-369-493-4917	Sequence 4917, Ap
56	41	39.4	478	15	US-10-369-493-7675	Sequence 7675, Ap
57	41	39.4	484	12	US-10-282-122A-75455	Sequence 75455, A
58	41	39.4	484	14	US-10-156-761-12673	Sequence 12673, A
59	41	39.4	498	15	US-10-369-493-15858	Sequence 15858, A
60	41	39.4	512	12	US-10-282-122A-49407	Sequence 49407, A
61	41	39.4	597	14	US-10-269-017-18	Sequence 18, Appl
62	41	39.4	1064	15	US-10-104-047-3830	Sequence 3830, Ap
63	40.5	38.9	482	14	US-10-156-761-11098	Sequence 11098, A
64	40.5	38.9	745	15	US-10-017-161-1972	Sequence 1972, Ap
65	40.5	38.9	745	15	US-10-292-798-1620	Sequence 1620, Ap
66	40	38.5	122	12	US-10-424-599-265930	Sequence 265930,
67	40	38.5	131	12	US-10-424-599-249049	Sequence 249049,
68	40	38.5	235	12	US-10-425-114-65890	Sequence 65890, A
69	40	38.5	296	12	US-10-425-114-42184	Sequence 42184, A
70	40	38.5	320	9	US-09-738-626-6358	Sequence 6258, Ap
71	40	38.5	346	14	US-10-156-761-8608	Sequence 8608, Ap
72	40	38.5	359	9	US-09-815-242-5116	Sequence 5116, Ap
73	40	38.5	359	12	US-10-282-122A-43489	Sequence 43489, A
74	40	38.5	363	12	US-10-424-599-282312	Sequence 282312,
75	40	38.5	370	12	US-10-425-114-60241	Sequence 60241, A
76	40	38.5	372	12	US-10-282-122A-66951	Sequence 66951, A
77	40	38.5	448	12	US-10-425-114-45312	Sequence 45312, A
78	40	38.5	452	12	US-10-424-599-177415	Sequence 177415,
79	40	38.5	452	12	US-10-425-114-63993	Sequence 63993, A
80	40	38.5	484	12	US-10-282-122A-56323	Sequence 56323, A
81	40	38.5	490	15	US-10-369-493-15077	Sequence 15077, A
82	40	38.5	499	15	US-10-369-493-11671	Sequence 11671, A
83	40	38.5	499	15	US-10-369-493-14468	Sequence 14468, A
84	40	38.5	514	12	US-10-412-699B-1026	Sequence 1026, Ap
85	40	38.5	514	15	US-10-374-780A-534	Sequence 534, Ap
86	40	38.5	582	15	US-10-369-493-19315	Sequence 19315, A
87	40	38.5	650	12	US-10-424-599-278827	Sequence 278827,
88	40	38.5	691	15	US-10-369-493-10173	Sequence 10173, A

Query Match 43.3%; Score 45; DB 12; Length 233;
Best Local Similarity 69.2%; Pred. No. 53;
Matches 9; Conservative 1; Mismatches 3; Indels

RESULT 8

US-10-282-122A-76545

Sequence 76545, Application US/10282122A

Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu

APPLICANT: Zamudio, Carlos

APPLICANT: Malone, Cheryl

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari

APPLICANT: Zyskind, Judith

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John

APPLICANT: Carr, Grant

APPLICANT: Yamamoto, Robert

APPLICANT: Forsyth, R.

APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

```
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 76545
; LENGTH: 287
; TYPE: PRT
; ORGANISM: Treponema pallidum
US-10-282-122A-76545

Query Match      43.3%; Score 45; DB 12; Length 287;
Best Local Similarity 36.8%; Pred. No. 67;
Matches 7; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY      2 QRAVRPALSGFGRVSGC 20
DB      186 KKSAGFIGGFAGSVGAGC 204
      : : : : : | | | | |
      : : : : : | | | | |

RESULT 9
US-10-369-493-11875
; Sequence 11875, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 11875
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Mesorhizobium loti
US-10-369-493-11875

Query Match      43.3%; Score 45; DB 15; Length 299;
Best Local Similarity 44.4%; Pred. No. 69;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY      1 LQRAVRPALSGFGRVGS 18
DB      5 LNRSTRVIVGGTGRKGS 22
      | | | : | | | | |
      | | | : | | | | |

RESULT 10
US-09-815-242-10383
; Sequence 10383, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; PROKARYOTES
```

```
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10383
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10383

Query Match      43.3%; Score 45; DB 9; Length 484;
Best Local Similarity 47.4%; Pred. No. 1,2e+02;
Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY      2 QRAVRPALSGFGRVSGC 20
DB      87 QRVLRPAILWNGRCQAQEC 105
      | | | | | | | | |
      | | | | | | | | |

RESULT 11
US-10-282-122A-56739
; Sequence 56739, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
```

; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 56739
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-282-122A-56739

Query Match 43.3%; Score 45; DB 12; Length 484;
Best Local Similarity 47.4%; Pred. No. 1.2e+02;
Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 2 QRAVRPALSGFGRVSGC 20
||:||||:||||
DB 87 QRVLRPAILWNGRCQAEC 105

RESULT 12
US-10-369-493-23598
; Sequence 23598, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 23598
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-369-493-23598

Query Match 43.3%; Score 45; DB 15; Length 484;
Best Local Similarity 47.4%; Pred. No. 1.2e+02;
Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 2 QRAVRPALSGFGRVSGC 20
||:||||:||||
DB 87 QRVLRPAILWNGRCQAEC 105

RESULT 13
US-09-864-408A-6820
; Sequence 6820, Application US/09864408A
; Publication No. US2004000947A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shimkets, Richard A.
; TITLE OF INVENTION: No. US2004000947A1 Human Polynucleotides and Polypeptides Encod
; FILE REFERENCE: 21402-012
; CURRENT APPLICATION NUMBER: US/09/864,408A
; CURRENT FILING DATE: 2001-05-24
; PRIOR APPLICATION NUMBER: 60/206,690
; PRIOR FILING DATE: 2000-05-24
; NUMBER OF SEQ ID NOS: 9068
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6820
; LENGTH: 73
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-864-408A-6820

Query Match 42.8%; Score 44.5; DB 11; Length 73;
Best Local Similarity 44.4%; Pred. No. 19;
Matches 12; Conservative 4; Mismatches 4; Indels 7; Gaps 2;

QY 1 LQRAVRPALSGFGRVSGC 20
||:||||:||||
DB 13 LVRSIVPALTSKHKRGDGRIGVGC 39

RESULT 14
US-10-424-599-150057
; Sequence 150057, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 150057
; LENGTH: 99
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(99)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_106522C.1.pep
US-10-424-599-150057

Query Match 42.8%; Score 44.5; DB 12; Length 99;
Best Local Similarity 64.7%; Pred. No. 26;
Matches 11; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 4 AVRPAALSGFGRVSGC 19
||:||||:||||
DB 46 SVPPVLSFFDGRLLGSG 62

RESULT 15
US-10-369-493-1228
; Sequence 1228, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1228
; LENGTH: 301
; TYPE: PRT
; ORGANISM: Methanobacterium thermoautotrophicum
US-10-369-493-1228

Query Match 42.8%; Score 44.5; DB 15; Length 301;
Best Local Similarity 50.0%; Pred. No. 84;
Matches 10; Conservative 2; Mismatches 5; Indels 3; Gaps 1;

QY 3 RAVRPALSGDGR---VSGS 19
Db 133 QTIREKLGDFDGLRVFVGG 152

RESULT 16
US-10-424-599-177904
; Sequence 177904, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 177904
; LENGTH: 133
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(133)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_131663C.1.pep
US-10-424-599-177904

Query Match 42.3%; Score 44; DB 12; Length 133;
Best Local Similarity 53.3%; Pred. No. 43;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 LORAVRPALSGDGR 15
Db 22 LRSVRPRLANYQGR 36

RESULT 17
US-10-369-493-3741
; Sequence 3741, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 3741
; LENGTH: 184
; TYPE: PRT
; ORGANISM: Neurospora crassa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(184)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-3741

Query Match 42.3%; Score 44; DB 15; Length 184;
Best Local Similarity 47.4%; Pred. No. 60;
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 2 QRAVRPALSGDGRVSGC 20
Db 163 EAATRAALLTFEKEGSGC 181

RESULT 18
US-10-424-599-198861
; Sequence 198861, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 198861
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_21596C.1.pep
US-10-424-599-198861

Query Match 42.3%; Score 44; DB 12; Length 364;
Best Local Similarity 61.5%; Pred. No. 1.2e+02;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 QRAVRPALSGFDG 14
Db 188 QRAARPAARGYNG 200

RESULT 19
US-10-424-599-241830
; Sequence 241830, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 241830
; LENGTH: 111
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(111)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_603C.1.pep
US-10-424-599-241830

Query Match 41.8%; Score 43.5; DB 12; Length 111;
Best Local Similarity 50.0%; Pred. No. 42;
Matches 11; Conservative 2; Mismatches 6; Indels 3; Gaps 1;

QY 2 QRAVRPALSG---FDGRVSGC 20
Db 28 QRRVRPTLSAEXIYDGLSLRC 49

```
RESULT 20
US-10-425-114-46577
; Sequence 46577, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 46577
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700445211_FLI.pep
US-10-425-114-46577

Query Match 41.3%; Score 43; DB 12; Length 179;
Best Local Similarity 50.0%; Pred. No. 84;
Matches 10; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 1 LQAVRPALSGDGRVSGC 20
DB 66 LHHAADPASCGRDGGGIGC 85

RESULT 21
US-10-156-761-10036
; Sequence 10036, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR FILING DATE: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10036
; LENGTH: 298
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10036

Query Match 41.3%; Score 43; DB 14; Length 298;
Best Local Similarity 61.5%; Pred. No. 1.4e+02;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 6 RPALSGDGRVGS 18
DB 278 RPLVGGEDGRTGA 290

RESULT 22
US-10-425-114-44283
; Sequence 44283, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 44283
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 701162958_FLI.pep
US-10-425-114-44283

Query Match 41.3%; Score 43; DB 12; Length 299;
Best Local Similarity 50.0%; Pred. No. 1.4e+02;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 LQAVRPALSGDGRV 16
DB 229 LQATSPAVQAFGRV 244

RESULT 23
US-10-015-115-14
; Sequence 14, Application US/10015115
; Publication No. US20030207800A1
; GENERAL INFORMATION:
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Zerhusen, Bryan D
; APPLICANT: Patturajan, Meera
; APPLICANT: Guo, Xiaojia
; APPLICANT: Kekuda, Ramesha
; APPLICANT: Gangolli, Esha A
; APPLICANT: Shinkets, Richard A
; APPLICANT: Taupier, Raymond J
; APPLICANT: Li, Li
; APPLICANT: Padigaru, Muralidhara
; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
; FILE OF INVENTION: Using the Same
; FILE REFERENCE: 21402-211
; CURRENT APPLICATION NUMBER: US/10/015,115
; CURRENT FILING DATE: 2002-09-23
; PRIOR APPLICATION NUMBER: 60/248,153
; PRIOR FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: 60/249,598
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/264,240
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/266,127
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/269,562
; PRIOR FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: 60/304,348
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/309,261
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/313,283
; NUMBER OF SEQ ID NOS: 205
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 14
```

```

; LENGTH: 2695
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-015-115-14

Query Match 41.3%; Score 43; DB 15; Length 2695;
Best Local Similarity 58.3%; Pred. No. 1.4e+03;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 VRPALSQFDGRV 16
Db 2583 LRPALGLEGRI 2594

RESULT 24
US-10-015-115-16
; Sequence 16, Application US/10015115
; Publication No. US20030207800A1
; GENERAL INFORMATION:
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Zerhusen, Bryan D
; APPLICANT: Patturajan, Meera
; APPLICANT: Guo, Xiaojia
; APPLICANT: Kekuda, Ramesha
; APPLICANT: Gangolli, Esha A
; APPLICANT: Shimkets, Richard A
; APPLICANT: Taupier, Raymond J
; APPLICANT: Li, Li
; APPLICANT: Padigar, Muralidhara
; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
; FILE REFERENCE: 21402-211
; CURRENT APPLICATION NUMBER: US/10/015,115
; CURRENT FILING DATE: 2002-09-23
; PRIOR FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: 60/248,153
; PRIOR FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: 60/249,598
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/264,240
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/266,127
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/269,562
; PRIOR FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: 60/304,348
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/309,261
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/313,283
; PRIOR FILING DATE: 2001-08-17
; NUMBER OF SEQ ID NOS: 205
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 2844
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-015-115-5

Query Match 41.3%; Score 43; DB 15; Length 2844;
Best Local Similarity 58.3%; Pred. No. 1.5e+03;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 VRPALSQFDGRV 16
Db 2732 LRPALGLEGRI 2743

RESULT 26
US-10-015-115-12
; Sequence 12, Application US/10015115
; Publication No. US20030207800A1
; GENERAL INFORMATION:
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Zerhusen, Bryan D
; APPLICANT: Patturajan, Meera
; APPLICANT: Guo, Xiaojia
; APPLICANT: Kekuda, Ramesha
; APPLICANT: Gangolli, Esha A
; APPLICANT: Shimkets, Richard A
; APPLICANT: Taupier, Raymond J
; APPLICANT: Li, Li
; APPLICANT: Padigar, Muralidhara
; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
; FILE REFERENCE: 21402-211
; CURRENT APPLICATION NUMBER: US/10/015,115
; CURRENT FILING DATE: 2002-09-23
; PRIOR FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: 60/248,153
; PRIOR FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: 60/249,598
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/264,240
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/266,127
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/269,562
; PRIOR FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: 60/304,348
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/309,261
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/313,283
; PRIOR FILING DATE: 2001-08-17
; NUMBER OF SEQ ID NOS: 205
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 2757
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-015-115-16

Query Match 41.3%; Score 43; DB 15; Length 2757;
Best Local Similarity 58.3%; Pred. No. 1.5e+03;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 VRPALSQFDGRV 16
Db 2645 LRPALGLEGRI 2656

RESULT 25
US-10-015-115-8
; Sequence 8, Application US/10015115
; Publication No. US20030207800A1

```



```

; PRIOR APPLICATION NUMBER: 60/248,153
; PRIOR FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: 60/249,598
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/264,240
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/266,127
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/269,562
; PRIOR FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: 60/304,348
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/309,261
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/313,283
; PRIOR FILING DATE: 2001-08-17
; NUMBER OF SEQ ID NOS: 205
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 2845
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-015-115-12

Query Match 41.3%; Score 43; DB 15; Length 2845;
Best Local Similarity 58.3%; Pred. NO. 1.5e+03;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 VRPALSQPDGRV 16
Db 2733 LRPALGLEGRI 2744

RESULT 27
US-10-015-115-10
; Sequence 10, Application US/10015115
; Publication No. US20030207800A1
; GENERAL INFORMATION:
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Zehrusen, Bryan D
; APPLICANT: Patturajan, Meera
; APPLICANT: Guo, Xiaojia
; APPLICANT: Gangolli, Esha A
; APPLICANT: Shinkets, Richard A
; APPLICANT: Taupier, Raymond J
; APPLICANT: Li, Li
; APPLICANT: Padigaru, Muralidhara
; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
; FILE REFERENCE: 21402-211
; CURRENT APPLICATION NUMBER: US/10/015,115
; CURRENT FILING DATE: 2002-09-23
; PRIOR APPLICATION NUMBER: 60/248,153
; PRIOR FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: 60/249,598
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/264,240
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/266,127
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/269,562
; PRIOR FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: 60/304,348
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/309,261
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/313,283
; PRIOR FILING DATE: 2001-08-17
; NUMBER OF SEQ ID NOS: 205
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 2995
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: VARIANT
; LOCATION: (49)
; OTHER INFORMATION: Wherein Xaa is any amino acid as defined in the
; OTHER INFORMATION: specification.
; NAME/KEY: VARIANT
; LOCATION: (98)
; OTHER INFORMATION: Wherein Xaa is any amino acid as defined in the
; OTHER INFORMATION: specification.
; NAME/KEY: VARIANT
; LOCATION: (104)
; OTHER INFORMATION: Wherein Xaa is any amino acid as defined in the
; OTHER INFORMATION: specification.

```

OTHER INFORMATION: specification.
 FEATURE:
 NAME/KEY: VARIANT
 LOCATION: (106)
 OTHER INFORMATION: Wherein Xaa is any amino acid as defined in the
 OTHER INFORMATION: specification.
 US-10-015-115-6

Query Match 41.3%; Score 43; DB 15; Length 2995;
 Best Local Similarity 58.3%; Pred. No. 1.6e+03;
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 VRPALSGFDGRV 16
 :|||:|:
 DB 2883 LRPALGLEGRI 2894

RESULT 29
 US-10-424-599-275944
 ; Sequence 275944, Application US/10424599
 ; Publication No. US20040031072A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa Thomas J
 ; APPLICANT: Kovalic David K
 ; APPLICANT: Zhou Yihua
 ; APPLICANT: Cao Yongwei
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53223)B
 ; CURRENT APPLICATION NUMBER: US/10/424,599
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 285684
 ; SEQ ID NO 275944
 ; LENGTH: 87
 ; TYPE: PRT
 ; ORGANISM: Glycine max
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT3847_91198C.1.psp
 US-10-424-599-275944

Query Match 40.9%; Score 42.5; DB 12; Length 87;
 Best Local Similarity 52.2%; Pred. No. 47;
 Matches 12; Conservative 2; Mismatches 4; Indels 5; Gaps 2;

QY 1 LQRAVRPALSG---FDGRVSGSC 20
 |:|||:|:
 DB 35 LDRIRPAQAGGLNFVGR--SGC 55

RESULT 30
 US-10-369-493-19465
 ; Sequence 19465, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 19465
 ; LENGTH: 543
 ; TYPE: PRT
 ; ORGANISM: Myxococcus xanthus
 US-10-369-493-19465

Query Match 40.9%; Score 42.5; DB 15; Length 543;
 Best Local Similarity 55.0%; Pred. No. 3.2e+02;
 Matches 11; Conservative 2; Mismatches 6; Indels 1; Gaps 1;
 QY 1 LQRAVRPALSGFDGRV-GSG 19
 |:|||:|:
 DB 315 LERAERPGRAGLRGRPRGSG 334
 Search completed: May 13, 2004, 07:17:02
 Job time : 214 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2004, 06:49:47 ; Search time 12 Seconds
(without alignments)
160.319 Million cell updates/sec

Title: 09549186-7

Perfect score: 104

Sequence: 1 LQRAVRPALSGFGRVSGC 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

PIR_78:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	46.2	376	2 C75580	adenine deaminase-
2	46	44.2	293	2 F70724	hypothetical prote
3	46	44.2	1112	2 S70522	cyclic nucleotide
4	45	43.3	287	2 D71304	probable phosphati
5	45	43.3	484	1 KIECKY	xylokkinase (SC 2
6	45	43.3	484	2 E86031	xylokkinase (impo
7	45	43.3	484	2 G91184	xylokkinase (impo
8	44.5	42.8	301	2 E69059	ornithine carbamoy
9	44	42.3	136	2 E95333	hypothetical prote
10	44	42.3	477	2 H84222	hypothetical prote
11	44	42.3	940	2 S49087	lactoferrin bindin
12	44	42.3	943	2 G81070	lactoferrin-bindin
13	44	42.3	944	2 C81798	lactoferrin bindin
14	44	42.3	1355	1 VGBE11	149K glycoprotein
15	44	42.3	1746	2 D83181	hypothetical prote
16	44	42.3	1964	2 T09059	notch4 - mouse
17	43.5	41.8	191	2 A65137	hypothetical 21.0
18	43.5	41.8	191	2 H91160	hypothetical prote
19	43.5	41.8	191	2 G86006	hypothetical prote
20	43.5	41.8	191	2 AC0016	conserved hypotet
21	43.5	41.8	191	2 AC0997	conserved hypotet
22	43	41.3	1747	2 A54121	collagen alpha-4 c
23	42	40.4	106	2 AF2742	ferrodoxin [import
24	42	40.4	135	2 AB2983	conserved hypotet
25	42	40.4	135	2 E98300	hypothetical prote
26	42	40.4	143	2 E97523	hypothetical prote
27	42	40.4	176	2 B72698	ferrodoxin, 2fe-2S
28	42	40.4	224	2 A70728	hypothetical prote
29	42	40.4	263	2 AC1313	dihydrodipicolinat

30	42	40.4	330	2 B87520	conserved hypotet
31	42	40.4	415	2 T23215	hypothetical prote
32	42	40.4	466	2 T45585	hypothetical prote
33	42	40.4	488	2 A95326	probable argininos
34	42	40.4	575	2 D69512	acetylactate synth
35	42	40.4	616	2 H87654	ABC transporter, H
36	41.5	39.9	101	2 A72557	hypothetical prote
37	41.5	39.9	283	2 D84868	probable endochiti
38	41.5	39.9	396	2 A71281	probable S-adenosy
39	41.5	39.9	465	1 WMB344	maturation protein
40	41.5	39.9	482	2 B82028	adhesin NMA0324 [1
41	41.5	39.9	621	1 S59632	endo-1,4-beta-xyla
42	41.5	39.9	986	2 G65116	hypothetical prote
43	41	39.4	55	2 H97591	hypothetical prote
44	41	39.4	135	2 S11478	circular genome pr
45	41	39.4	273	2 H69206	dihydrodipicolinat
46	41	39.4	282	2 A31765	heterogeneous ribo
47	41	39.4	296	2 T13885	NADH2 dehydrogenas
48	41	39.4	330	2 S47491	genome polyprotein
49	41	39.4	361	2 G70959	hypothetical prote
50	41	39.4	484	2 AD0980	xylokkinase (EC 2
51	41	39.4	494	2 B95411	probable aldehyde
52	41	39.4	552	2 E75032	carbon starvation
53	41	39.4	734	2 AF3108	GGDEF family prote
54	41	39.4	747	2 F98178	nitrogen fixation
55	41	39.4	812	2 S31521	collagen COL1 - f
56	40.5	38.9	1266	2 A85989	hypothetical prote
57	40.5	38.9	1266	2 F91143	hypothetical prote
58	40	38.5	113	2 A13371	ferrodoxin, 2fe-2S
59	40	38.5	203	2 C75366	chromosome partiti
60	40	38.5	222	2 AH2619	conserved hypotet
61	40	38.5	222	2 H97401	probable endopepti
62	40	38.5	359	2 E83312	hypothetical prote
63	40	38.5	363	2 T36408	probable esterase
64	40	38.5	385	2 G90274	hypothetical prote
65	40	38.5	394	2 F72517	hypothetical prote
66	40	38.5	445	2 H96560	hypothetical prote
67	40	38.5	503	2 AB2938	alpha-L-arabinofur
68	40	38.5	503	2 E98344	hypothetical prote
69	40	38.5	892	2 T09071	SH3 domains-contai
70	40	38.5	1037	2 T13943	phospholipase D (E
71	40	38.5	1075	2 T46635	phospholipase D (E
72	40	38.5	1707	2 F96711	hypothetical prote
73	40	38.5	1707	2 S77908	hypothetical prote
74	40	38.5	2117	2 T36180	CD4 peptide synthe
75	40	38.5	2157	2 A13009	peptide synthetase
76	40	38.5	2566	2 E98274	hypothetical prote
77	39.5	38.0	195	2 B82040	conserved hypotet
78	39.5	38.0	335	2 F87651	glycerolaldehyde 3-p
79	39.5	38.0	335	2 AH3290	glycerolaldehyde-3-p
80	39.5	38.0	397	2 T08345	hypothetical prote
81	39.5	38.0	1344	2 T14316	rig-1 protein - mo
82	39	37.5	113	2 FC4289	brain and muscle A
83	39	37.5	159	2 B25297	chorion class B pr
84	39	37.5	177	2 S39859	transcription anti
85	39	37.5	177	2 AG1105	transcription anti
86	39	37.5	207	2 AB1877	general secretion
87	39	37.5	263	2 AC1685	dihydrodipicolinat
88	39	37.5	275	2 T11810	ribosomal protein
89	39	37.5	276	2 T07531	ribosomal protein
90	39	37.5	285	2 A71553	hypothetical prote
91	39	37.5	290	2 A26885	heterogeneous nucl
92	39	37.5	293	2 E89796	N-acetylneuraminat
93	39	37.5	303	2 C34504	heterogeneous ribo
94	39	37.5	311	2 T43947	N-acetyl-gamma-glu
95	39	37.5	329	2 E72618	hypothetical prote
96	39	37.5	366	2 F70618	probable pntAA pro
97	39	37.5	367	2 B87239	pyridine transhydr
98	39	37.5	381	2 AB3297	ABC transporter AT
99	39	37.5	382	2 T03568	probable malate sy
100	39	37.5	387	2 H88012	protein K1084.2 [1

ALIGNMENTS

```
RESULT 1
C75580
adenine deaminase-related protein - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: C75580
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
; M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896; PMID:10567266
A;Accession: C75580
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-376 <WHI>
A;Cross-references: GB:AE001863; GB:AE001825; NID:g6460670; PIDN:AAF12376.1; PID:g646067
A;Experimental source: strain R1
C;Genetics:
A;Gene: DRA0268
A;Map position: 2

Query Match 46.2%; Score 48; DB 2; Length 376;
Best Local Similarity 64.7%; Pred. No. 8;
Matches 11; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 RAVRPALSGFDGRVSG 19
||| ||| ||| ||| |||
Db 275 RAVAPALRGSDRPASG 291

RESULT 2
F70724
hypothetical protein Rv2575 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: F70724
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.
; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: F70724
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-293 <COL>
A;Cross-references: GB:277724; GB:AL123456; NID:g3261620; PIDN:CAB01268.1; PID:g1478237
A;Experimental source: strain H37RV
C;Genetics:
A;Gene: Rv2575

Query Match 44.2%; Score 46; DB 2; Length 293;
Best Local Similarity 53.3%; Pred. No. 13;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 6 RPALSGFDGRVSG 20
||| : ||| ||| |||
Db 105 RPHRLFGQVGIGC 119

RESULT 3
S70522
cyclic nucleotide phosphodiesterase, cGMP-inhibited (EC 3.1.4.-) - human
C;Species: Homo sapiens (man)
C;Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 19-May-2000
C;Accession: S70522
R;Murata, T.; Taira, M.; Manganiello, V.C.
FEBS Lett. 390, 29-33, 1996
```

```
A;Title: Differential expression of cGMP-inhibited cyclic nucleotide phosphodiesterases i
A;Reference number: S70522; MUID:96314543; PMID:8706823
A;Accession: S70522
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: mRNA
A;Residues: 1-1112 <MUR>
A;Cross-references: EMBL:U38178; NID:g1145301; PIDN:AAC50724.1; PID:g1145302
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1996
C;Superfamily: cyclic-nucleotide phosphodiesterase, cGMP-inhibited; 3',5'-cyclic-nucleot
C;Keywords: phosphoric diester hydrolase
F;736-1006/Domain: 3',5'-cyclic-nucleotide phosphodiesterase homology <CNP>

Query Match 44.2%; Score 46; DB 2; Length 1112;
Best Local Similarity 55.0%; Pred. No. 49;
Matches 11; Conservative 2; Mismatches 5; Indels 2; Gaps 1;

QY 1 LQAVRPALSGFDGRVSGC 20
| : ||| ||| | : |||
Db 240 LPSALRPLSLGLVG--GAGC 257

RESULT 4
D71304
probable phosphatidate cytidyltransferase (cdsa) - syphilis spirochete
C;Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C;Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 17-Mar-2000
C;Accession: D71304
R;Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDo
rthy, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A;Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A;Reference number: A71250; MUID:98332770; PMID:9665876
A;Accession: D71304
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-287 <COL>
A;Cross-references: GB:AE001235; GB:AE000520; NID:g3322893; PIDN:AAC65575.1; PID:g332289
A;Experimental source: strain Nichols
C;Genetics:
A;Gene: TP0602
C;Superfamily: phosphatidate cytidyltransferase

Query Match 43.3%; Score 45; DB 2; Length 287;
Best Local Similarity 36.8%; Pred. No. 18;
Matches 7; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 2 QRVRPALSGFDGRVSGC 20
::: : ||| ||| |||
Db 186 KXSIAFGIFGFGAGVGAGC 204

RESULT 5
KIECXY
xylulokinase (EC 2.7.1.17) - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change 01-Mar-2002
C;Accession: A30266; S47785; F65155
R;Lawlis, V.B.; Dennis, M.S.; Chen, E.Y.; Smith, D.H.; Henner, D.J.
Appl. Environ. Microbiol. 47, 15-21, 1984
A;Title: Cloning and sequencing of the xylulose isomerase and xylulose kinase genes of Esc
A;Reference number: A90043; MUID:84126725; PMID:6320721
A;Accession: A30266
A;Molecule type: DNA
A;Residues: 1-484 <LAW>
A;Cross-references: GB:K01996; NID:g148278; PIDN:AAA24769.1; PID:g148280
R;Plunkett, G.
submitted to the EMBL Data Library, March 1994
A;Reference number: S47666
A;Accession: S47785
A;Molecule type: DNA
A;Residues: 1-484 <PLU>
A;Cross-references: EMBL:U00039; NID:g466582; PIDN:AAB18541.1; PID:g466702
```

R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of *Escherichia coli* K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: E65155
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-484 <BLAT>
A;Cross-references: GB:AE000433; GB:U00096; NID:gl789977; PIDN:AAC76588.1; PID:gl789987;
A;Experimental source: strain K-12, substrain MG1655
A;Comment: This enzyme catalyzes the phosphorylation of xylulose to xylulose-5-phosphate
C;Genetics:
A;Gene: xylB
A;Map position: 80 min
C;Superfamily: xylulokinase
C;Keywords: phosphotransferase

Query Match 43.3%; Score 45; DB 1; Length 484;
Best Local Similarity 47.4%; Pred. No. 31;
Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 2 QRAVRPALSGFDCRGVSGC 20
||:||||:||||
Db 87 QRVLRPAILLWNGRCQAQEC 105

RESULT 6
E86031
xylulokinase [imported] - *Escherichia coli* (strain O157:H7, substrain EDL933)
C;Species: *Escherichia coli*
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 02-Nov-2001
C;Accession: E86031
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimailanta, B.; Potamouisis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: E86031
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-484 <STO>
A;Cross-references: GB:AE005174; NID:gl2518302; PIDN:AGS8713.1; GSPDB:GN00145; UWGP:Z49
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: xylB
C;Superfamily: xylulokinase

Query Match 43.3%; Score 45; DB 2; Length 484;
Best Local Similarity 47.4%; Pred. No. 31;
Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 2 QRAVRPALSGFDCRGVSGC 20
||:||||:||||
Db 87 QRVLRPAILLWNGRCQAQEC 105

RESULT 7
G91184
xylulokinase [imported] - *Escherichia coli* (strain O157:H7, substrain RIMD 0509952)
C;Species: *Escherichia coli*
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 02-Nov-2001
C;Accession: G91184
R;Hayaishi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gene
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: G91184
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-484 <HAY>
A;Cross-references: GB:BA000007; PIDN:BA837870.1; PID:gl33363921; GSPDB:GN00154

A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECG4447
C;Superfamily: xylulokinase

Query Match 43.3%; Score 45; DB 2; Length 484;
Best Local Similarity 47.4%; Pred. No. 31;
Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 2 QRAVRPALSGFDCRGVSGC 20
||:||||:||||
Db 87 QRVLRPAILLWNGRCQAQEC 105

RESULT 8
E69059
ornithine carbamoyltransferase - *Methanobacterium thermoautotrophicum* (strain Delta H)
C;Species: *Methanobacterium thermoautotrophicum*
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 18-Jun-1999
C;Accession: E69059
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of *Methanobacterium thermoautotrophicum* Delta H: functi
A;Reference number: A69000; MUID:98037514; PMID:9371463
A;Accession: E69059
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-301 <MTH>
A;Cross-references: GB:AE000906; GB:AE000666; NID:g2622557; PIDN:AAB85921.1; PID:g2622557;
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH1446
C;Superfamily: ornithine carbamoyltransferase; aspartate/ornithine carbamoyltransferase
F;2-298/Domain: aspartate/ornithine carbamoyltransferase homology <ACT>

Query Match 42.8%; Score 44.5; DB 2; Length 301;
Best Local Similarity 50.0%; Pred. No. 23;
Matches 10; Conservative 2; Mismatches 5; Indels 3; Gaps 1;

QY 3 RAVRPALSGFDCRGVSGC 19
:|:||||:||||
Db 133 QTIREKLGFGDRLVFGDG 152

RESULT 9
B95333
hypothetical protein Sma1053 [imported] - *Sinorhizobium meliloti* (strain 1021) magaplasma
C;Species: *Sinorhizobium meliloti*
C;Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C;Accession: B95333
R;Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows
; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yen, K.C.
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A;Title: Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti*
A;Reference number: A95262; MUID:21396509; PMID:11481432
A;Accession: B95333
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-136 <KUR>
A;Cross-references: GB:AE006469; PIDN:AAK65228.1; PID:gl4523677; GSPDB:GN00165
A;Experimental source: strain 1021, megaplasmid pSymA
R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, P.; Barloy-Hubler,
Pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A;Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
A;Reference number: A96039; MUID:21368234; PMID:11474104
A;Contents: annotation
C;Genetics:

A;Gene: Sma1053
A;Genome: plasmid

Query Match 42.3%; Score 44; DB 2; Length 136;
Best Local Similarity 66.7%; Pred. No. 13;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 9 LSGFGDGRVSGC 20
| | | | | | | | | |
Db 86 LVFGSGRMGGC 97

RESULT 10
H84222
hypothetical protein Vng0650c [imported] - Halobacterium sp. NRC-1
C;Species: Halobacterium sp. NRC-1
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C;Accession: H84222
R;Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A;Title: Genome sequence of Halobacterium species NRC-1.
A;Reference number: A84160; MUID:20504483; PMID:11016950
A;Accession: H84222
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-477 <STO>
A;Cross-references: GB:AE004437; NID:g10580238; PIDN:AAG19148.1; GSPDB:GN00138
C;Genetics:
A;Gene: VNG0650C
C;Superfamily: hypothetical protein MJ1633

Query Match 42.3%; Score 44; DB 2; Length 477;
Best Local Similarity 53.3%; Pred. No. 44;
Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 3 RAVRPALSGDGRVG 17
: | | | | | | | | | |
Db 141 QALRQTLQGDGRLG 155

RESULT 11
S49087
lactoferrin binding protein - Neisseria meningitidis
C;Species: Neisseria meningitidis
C;Date: 16-Feb-1995 #sequence_revision 12-May-1995 #text_change 22-Oct-1999
C;Accession: S49087
R;Pettersson, A.M.; Klarenbeek, X.Y.Z.; van Deurzen, X.Y.Z.; Poolman, X.Y.Z.; Tommassen,
submitted to the EMBL Data Library, June 1994
A;Description: Molecular characterization of the structural gene for the lacto-ferrin re
A;Reference number: S49087
A;Accession: S49087
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-940 <PET>
A;Cross-references: EMBL:X79838; NID:g509053; PIDN:CAA56233.1; PID:g509054
C;Superfamily: ferrichrome-iron receptor; tonB-dependent receptor amino-terminal homolog
P;73-215/Domain: tonB-dependent receptor amino-terminal homology <TNN>
P;609-940/Domain: tonB-dependent receptor carboxyl-terminal homology <TNC>

Query Match 42.3%; Score 44; DB 2; Length 940;
Best Local Similarity 60.0%; Pred. No. 85;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 5 VRPALSGDGRVSG 19
: | | | | | | | | | |
Db 122 VQSLSGYGGRGSG 136

RESULT 12
G81070

lactoferrin-binding protein A NMB1540 [imported] - Neisseria meningitidis (strain MC58 se
C;Species: Neisseria meningitidis
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C;Accession: G81070
R;Tetzelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.;
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizzi, M.
Science 287, 1809-1815, 2000
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ver
A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A;Reference number: A81000; MUID:20175755; PMID:10710307
A;Accession: G81070
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-943 <TET>
A;Cross-references: GB:AE002504; GB:AE002098; NID:g7226785; PIDN:AAF41895.1; PID:g7226785
C;Experimental source: serogroup B, strain MC58
C;Genetics:
A;Gene: NMB1540
C;Superfamily: ferrichrome-iron receptor; tonB-dependent receptor amino-terminal homology

Query Match 42.3%; Score 44; DB 2; Length 943;
Best Local Similarity 60.0%; Pred. No. 86;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 5 VRPALSGDGRVSG 19
: | | | | | | | | | |
Db 126 VQSLSGYGGRGSG 140

RESULT 13
C81798
lactoferrin binding protein A NMA1739 [imported] - Neisseria meningitidis (strain Z2491 s
C;Species: Neisseria meningitidis
C;Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C;Accession: C81798
R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morell
; Holroyd, S.; Jagers, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A;Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A;Reference number: A81775; MUID:20222556; PMID:10761919
A;Accession: C81798
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-944 <PAR>
A;Cross-references: GB:AL162757; GB:AL157959; NID:g7380371; PIDN:CAB84967.1; PID:g7380383
C;Experimental source: serogroup A, strain Z2491
C;Genetics:
A;Gene: lbpA; NMA1739
C;Superfamily: ferrichrome-iron receptor; tonB-dependent receptor amino-terminal homology

Query Match 42.3%; Score 44; DB 2; Length 944;
Best Local Similarity 60.0%; Pred. No. 86;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 5 VRPALSGDGRVSG 19
: | | | | | | | | | |
Db 126 VQSLSGYGGRGSG 140

RESULT 14
VGBE11
149K glycoprotein - ictalurid herpesvirus 1 (strain auburn 1)
C;Species: ictalurid herpesvirus 1
A;Title: host ictalurus punctatus (channel catfish)
C;Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 16-Jul-1999
C;Accession: B36791
R;Davison, A.J.
submitted to GenBank, January 1992
A;Description: Channel catfish virus: a new type of herpesvirus.
A;Reference number: A36804
A;Accession: B36791
A;Molecule type: DNA

A;Residues: 1-1355 <DAV>
A;Cross-references: GB:M75136; NID:G331209; PIDN:AAA89149.1; PID:G331256
R;Davison, A.J.
Virology 186, 9-14, 1992
A;Title: Channel catfish virus: a new type of herpesvirus.
A;Reference number: A39447; MUID:92087490; PMID:1727613
A;Contents: annotation
A;Note: neither amino acid nor nucleotide sequence is given
C;Genetics:
A;Gene: 46
C;Superfamily: ictaluriid herpesvirus 149K glycoprotein
C;Keywords: Glycoprotein
F:81,112,129,169,173,192,542,655,682,744,780,811,815,860,865,868,882,895,1195,1213,1225,
Query Match 42.3%; Score 44; DB 1; Length 1355;
Best Local Similarity 62.5%; Pred. No. 1.2e+02;
Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
QY 3 RAVRPALSGFGRVGS 18
Db 256 RADRPVSTGVDCRVGS 271
RESULT 15
D83181
hypothetical protein PA3728 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: D83181
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Boman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen.
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: D83181
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1746 <STO>
A;Cross-references: GB:AE004791; GB:AE004091; NID:99949882; PIDN:AAG07115.1; GSPDB:GN0015
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA3728
Query Match 42.3%; Score 44; DB 2; Length 1746;
Best Local Similarity 60.8%; Pred. No. 1.6e+02;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 4 AVRPAALSGFGRVGS 18
Db 429 ARQPARSGFLGRIGN 443
RESULT 16
T09059
notch4 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 08-Sep-2002
C;Accession: T09059
R;Rowen, L.; Mahairas, G.; Qin, S.; Ahearn, M.E.; Dankers, C.; Lasky, S.; Loretz, C.; Schmitt, J.; et al.
submitted to the EMBL Data Library, October 1997
A;Description: Sequence of the mouse major histocompatibility locus class III region.
A;Reference number: Z16543
A;Accession: T09059
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1964 <ROW>
A;Cross-references: EMBL:AF030001; NID:G2564945; PID:G2564947
C;Genetics:
A;Gene: notch4
A;Map position: 17
A;Introns: 22/1; 49/2; 148/1; 264/1; 305/1; 384/1; 436/1; 501/1; 539/1; 577/1; 618/1; 671/1; 1679/3; 1729/1; 1761/3

C;Superfamily: notch protein; ankyrin repeat homology; EGF homology
C;Keywords: receptor; signal transduction
F:514-545/Domain: EGF homology <EGF>
Query Match 42.3%; Score 44; DB 2; Length 1964;
Best Local Similarity 52.9%; Pred. No. 1.8e+02;
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;
QY 3 RAVRPALSGFGRVGS 19
Db 1165 RCQPCGASGCEGRGDS 1181
RESULT 17
A65137
hypothetical 21.0 kD protein in bioh-gntt intergenic region - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C;Accession: A65137
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, D.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: A65137
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-191 <BLAT>
A;Cross-references: GB:AE000417; GB:U00096; NID:G2367220; PIDN:AAC76439.1; PID:gi789819;
A;Experimental source: strain K-12, substrain MGL655
C;Genetics:
A;Gene: yhgI
Query Match 41.8%; Score 43.5; DB 2; Length 191;
Best Local Similarity 31.4%; Pred. No. 21;
Matches 11; Conservative 2; Mismatches 7; Indels 15; Gaps 1;
QY 1 LQAVRPALSGFGRV 20
Db 115 LQSQINPQLAGHGRVSLMEITDGYAILQFGGCG 149
RESULT 18
H91160
hypothetical protein ECs4256 [imported] - Escherichia coli (strain O157:H7, substrain RIM)
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: H91160
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic islands.
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: H91160
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-191 <HAY>
A;Cross-references: GB:BA000007; PIDN:BA837679.1; PID:gl3363730; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs4256
Query Match 41.8%; Score 43.5; DB 2; Length 191;
Best Local Similarity 31.4%; Pred. No. 21;
Matches 11; Conservative 2; Mismatches 7; Indels 15; Gaps 1;
QY 1 LQAVRPALSGFGRV 20
Db 115 LQSQINPQLAGHGRVSLMEITDGYAILQFGGCG 149
RESULT 19
G86006

Db 25 MENAVRNSVPGIDAECGGAC 44

RESULT 24
AB2983
conserved hypothetical protein Atu3467 [imported] - Agrobacterium tumefaciens (strain C58)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
R:Accession: AB2983
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monke, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; Karp, P.; Romero, P.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, S.E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AB2983
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-135 <KUR>
A:Cross-references: GB:AE008689; PIDN:AA144280.1; PID:gl7741866; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu3467
A:Map position: linear chromosome

Query Match 40.4%; Score 42; DB 2; Length 135;
Best Local Similarity 66.7%; Pred. No. 26;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 9 LSGFDGRVSGGC 20
||| :|||
Db 86 LVGFGTRMGSGC 97

RESULT 25
E98300
hypothetical protein AGR_L_2724 [imported] - Agrobacterium tumefaciens (strain C58, Cere)
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C:Accession: E98300
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: E98300
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-135 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK89927.1; PID:gl5159880; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L_2724
A:Map position: linear chromosome

Query Match 40.4%; Score 42; DB 2; Length 135;
Best Local Similarity 66.7%; Pred. No. 26;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 9 LSGFDGRVSGGC 20
||| :|||
Db 86 LVGFGTRMGSGC 97

RESULT 26
E97523
ferredoxin, 2Fe-2S (fdi) [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
R:Accession: E97523
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: E97523
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-143 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK87142.1; PID:gl5156412; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_2495
A:Map position: circular chromosome
C:Superfamily: ferredoxin [2Fe-2S]; ferredoxin [2Fe-2S] homology

Query Match 40.4%; Score 42; DB 2; Length 143;
Best Local Similarity 35.0%; Pred. No. 27;
Matches 7; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 1 LQRAVRPALSGFDGRVSGGC 20
::|||::|||
Db 62 MENAVRNSVPGIDAECGGAC 81

RESULT 27
B72698
hypothetical protein APE1002 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K.; DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix
A:Reference number: A72450; MUID:99310339; PMID:10382966
A:Accession: B72698
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-176 <KAW>
A:Cross-references: DDBJ:AP000060; MUID:95104189; PIDN:BAA79986.1; PID:d1043772; PID:gs5104
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE1002

Query Match 40.4%; Score 42; DB 2; Length 176;
Best Local Similarity 35.0%; Pred. No. 34;
Matches 7; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 1 LQRAVRPALSGFDGRVSGGC 20
::|||::|||
Db 4 VERLYKPRRQGLHGERGGGC 23

RESULT 28
A70728
hypothetical protein RV2557 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: A70728
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the Biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: A70728
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-224 <COL>
A:Cross-references: GB:Z77250; GB:AL123456; NID:G3261617; PIDN:CAB01047.1; PID:e255316; C:Genetics:
A:Gene: RV2557

```

Query Match      40.4%; Score 42; DB 2; Length 224;
Best Local Similarity 61.5%; Pred. No. 43;
Matches      8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY      5 VRPALSGFDGRVG 17
DB      44 VMPALQMDGICG 56

RESULT 29
AC1313
dihydrodipicolinate reductase homolog dapB [imported] - Listeria monocytogenes (strain B
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C:Accession: AC1313
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duesurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlant,
A:Title: Comparative Genomics of Listeria species.
A:Reference number: AB1077; PMID:21537279; PMID:11679669
A:Accession: AC1313
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-263 <GLA>
A:Cross-references: GB:NC_003210; PIDN:CAC99985.1; PID:gl6411360; GSPDB:GN00177
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: dapB
C:Superfamily: dihydrodipicolinate reductase

Query Match      40.4%; Score 42; DB 2; Length 263;
Best Local Similarity 61.5%; Pred. No. 50;
Matches      8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY      5 VRPALSGFDGRVG 17
DB      1 MRVAVSGFKGRMG 13

RESULT 30
B87520
conserved hypothetical protein CC2187 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: B87520
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: AB7249; PMID:21173698; PMID:11259647
A:Accession: B87520
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-330 <STO>
A:Cross-references: GB:AE005673; NID:gl3423688; PIDN:AAK24158.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC2187

Query Match      40.4%; Score 42; DB 2; Length 330;
Best Local Similarity 56.2%; Pred. No. 63;
Matches      9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY      1 LQAVRPALSGFDGRV 16
DB      268 MARAVRPCHTFDGDV 283

Search completed: May 13, 2004, 06:53:17
Job time : 17 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2004, 06:49:47 ; Search time 7.66667 Seconds
(without alignments)
135.835 Million cell updates/sec

Title: 09549186-7

Perfect score: 104

Sequence: 1 LQRAVRPALSGFDRVSGC 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB: ID	Description
1	46	44.2	293	1 YP75_MYCTU	Q50646 mycobacteri
2	46	44.2	1112	1 CN3B_HUMAN	Q13370 homo sapien
3	45	43.3	484	1 XYLX_ECOLI	P09099 escherichia
4	44.5	42.8	301	1 OTC_METTH	O27495 methanobact
5	44	42.3	405	1 IF4A_CRYPV	O02494 cryptospori
6	44	42.3	943	1 LBPA_NEIMB	O06379 neisseria m
7	44	42.3	944	1 LBPA_NEIMA	Q9JTK4 neisseria m
8	44	42.3	1355	1 VG46_HSV1	Q00104 ictaluriid h
9	44	42.3	1964	1 NTC4_MOUSE	P31695 mus musculu
10	43.5	41.8	191	1 YHGI_ECOLI	P46847 escherichia
11	42	40.4	148	1 REGQ_BPAPS	Q9T1U3 bacterioph
12	42	40.4	224	1 YP57_MYCTU	O50741 mycobacteri
13	42	40.4	263	1 DAPB_LISMO	Q8Y5Z6 listeria mo
14	42	40.4	415	1 BCAT_CABEL	P54688 caenorhabd
15	42	40.4	485	1 DPG2_HUMAN	Q9UHN1 homo sapien
16	42	40.4	488	1 ARL2_RHIME	Q92VM6 rhizobium m
17	41.5	39.9	396	1 METK_TREPA	O83772 treponema p
18	41.5	39.9	465	1 VP19_HSV14	P17586 herpes simp
19	41.5	39.9	986	1 YHDP_ECOLI	P46474 escherichia
20	41.5	39.9	1189	1 HAIR_HUMAN	Q43593 homo sapien
21	41	39.4	135	1 YOR2_COYMY	P19201 commelina y
22	41	39.4	273	1 DAPB_METTH	O26891 methanobact
23	41	39.4	282	1 ROC_XENLA	P19600 xenopus lae
24	41	39.4	953	1 SPB2_HUMAN	Q14151 homo sapien
25	40.5	38.9	969	1 ARVC_MOUSE	P98203 mus musculu
26	40	38.5	313	1 HXAB_MOUSE	P31311 mus musculu
27	40	38.5	448	1 SP52_HUMAN	Q99611 homo sapien
28	40	38.5	1200	1 ALA9_ARATH	O99X33 arabidopsis
29	40	38.5	1230	1 RPOM_HUMAN	Q00411 homo sapien
30	40	38.5	1302	1 ACSC_ACEXY	P37718 acetobacter
31	39	37.5	177	1 NUSG_BACSU	Q06795 bacillus su
32	39	37.5	183	1 NUSG_PASMU	Q9CK84 pasteurella
33	39	37.5	263	1 DAPB_LISIN	Q92AAL listeria in

34	39	37.5	1	1 RK2_PICAB	O62954 picea abies
35	39	37.5	1	1 RK2_PINTH	O62940 pinus thunb
36	39	37.5	1	1 NANA_STAAM	Q09WR1 staphylococ
37	39	37.5	1	1 NANA_STAAM	Q09WR1 staphylococ
38	39	37.5	1	1 ROC_MOUSE	Q07910 homo sapien
39	39	37.5	1	1 ROC_MOUSE	P02204 mus musculu
40	39	37.5	1	1 ARC2_THETH	O50146 thermus the
41	39	37.5	1	1 IF41_RABIT	P29562 oryctolagus
42	39	37.5	1	1 IF41_HUMAN	P29562 oryctolagus
43	39	37.5	1	1 CSIE_ECOLI	P54901 escherichia
44	39	37.5	1	1 APR2_ARATH	P54901 escherichia
45	39	37.5	1	1 APR3_ARATH	P54901 escherichia
46	39	37.5	1	1 APR1_ARATH	P52980 a 5'-adenyl
47	39	37.5	1	1 CYSN_BACTN	Q8AAP9 bacteroides
48	39	37.5	1	1 TWST_DROME	P10677 drosophila
49	39	37.5	1	1 ATPA_THIFE	P41167 thiobacillu
50	39	37.5	1	1 IF4B_HUMAN	P23588 homo sapien
51	39	37.5	1	1 IF2_STRCO	Q8CJQ8 streptomyce
52	39	37.5	1	1 IF2_STRAW	Q82K53 streptomyce
53	39	37.5	1	1 EMBC_MYCLE	Q9CDA7 mycobacteri
54	39	37.5	1	1 EMBC_MYCTU	P72059 mycobacteri
55	39	37.5	1	1 ALAB_ARATH	Q9SAF5 arabidopsis
56	39	37.5	1	1 RPA2_SCHPO	Q9P7X8 schizosacch
57	39	37.5	1	1 MPRI_BOVIN	P08169 bos taurus
58	39	37.5	1	1 TEGU_EBV	P03186 Epstein-bar
59	38.5	37.0	438	1 TEA4_CHICK	P48984 gallus gall
60	38	36.5	132	1 CHA2_BOMMO	P08825 bombyx mori
61	38	36.5	146	1 CDD_MOUSE	P56389 mus musculu
62	38	36.5	162	1 RRAA_PSESM	Q88346 pseudomonas
63	38	36.5	180	1 NUSG_ECOL6	Q919K0 escherichia
64	38	36.5	180	1 NUSG_ECOLI	P16921 escherichia
65	38	36.5	183	1 SODE_HABCO	P51547 haemochus
66	38	36.5	186	1 COAE_THETH	Q56416 thermus the
67	38	36.5	209	1 URK_CLOTE	Q89663 clostridium
68	38	36.5	243	1 PYRF_XANAC	Q8PE44 xanthomonas
69	38	36.5	243	1 PYRF_XANCP	Q893D7 xanthomonas
70	38	36.5	251	1 YMP3_STRCO	P43168 streptomyce
71	38	36.5	255	1 DAPB_STRMU	Q8DUL9 streptococc
72	38	36.5	329	1 COAA_STRCO	Q8D779 streptomyce
73	38	36.5	341	1 RTCA_PSEAE	Q9HVJ9 pseudomonas
74	38	36.5	353	1 DCAM_PEA	Q43820 pisum sativ
75	38	36.5	353	1 DCAM_VICFA	Q9M488 vicia faba
76	38	36.5	367	1 NOLF_RHIME	P25196 rhizobium m
77	38	36.5	407	1 IF42_HUMAN	Q14240 homo sapien
78	38	36.5	407	1 IF42_MOUSE	P10630 mus musculu
79	38	36.5	418	1 TRPB_SYNPX	O7T566 synchococc
80	38	36.5	440	1 CLPX_AZOVI	P33683 azotobacter
81	38	36.5	459	1 EX7L_YERPE	Q8ZCU2 yersinia pe
82	38	36.5	485	1 LEU2_ACTTI	P44427 actinoplan
83	38	36.5	527	1 NPPE_BACBR	P43263 bacillus br
84	38	36.5	543	1 RRP3_YEAST	P38712 saccharomyc
85	38	36.5	589	1 SPY_DROME	O44783 drosophila
86	38	36.5	598	1 CSTA_BACSU	P94532 bacillus su
87	38	36.5	598	1 YAAI_SCHPO	Q09795 schizosacch
88	38	36.5	609	1 GLCE_HUMAN	P54923 homo sapien
89	38	36.5	684	1 CDC4_CANAL	P53699 candida alb
90	38	36.5	736	1 DHB4_HUMAN	P51659 h peroxisom
91	38	36.5	834	1 FTSK_PSEPK	Q88F58 pseudomonas
92	38	36.5	900	1 IF38_ARATH	Q49160 arabidopsis
93	38	36.5	1036	1 PLDI_CRIGR	O08684 cricetus
94	38	36.5	1042	1 CORI_HUMAN	Q09545 homo sapien
95	38	36.5	1044	1 CARB_THIEVO	Q97A13 thermoplasm
96	38	36.5	1074	1 PLDI_HUMAN	Q13393 homo sapien
97	38	36.5	1074	1 PLDI_MOUSE	Q92280 mus musculu
98	38	36.5	1074	1 PLDI_RAT	P70436 rattus norv
99	38	36.5	1087	1 E4L3_HUMAN	Q9Y212 homo sapien
100	38	36.5	1189	1 YJH6_YEAST	P47035 saccharomyc

ALIGNMENTS

RESULT 1

DR TIGRFAMS; TIGR00658; orni carb tr; 1.
DR PROSITE; PS00097; CARBAMOYLTRANSFERASE; 1.
KW Arginine biosynthesis; Transferase; Complete proteome.
FT SITE 24 24 IMPORTANT FOR STRUCTURAL INTEGRITY (BY
FT SIMILARITY).
FT SITE 46 50 CARBAMOYLPHOSPHATE BINDING (BY
FT SIMILARITY).
FT SITE 97 97 CARBAMOYLPHOSPHATE BINDING (BY
FT SIMILARITY).
FT SITE 124 124 CARBAMOYLPHOSPHATE BINDING (BY
FT SIMILARITY).
FT SITE 137 137 IMPORTANT FOR STRUCTURAL INTEGRITY (BY
FT SIMILARITY).
FT SITE 257 260 ORNITHINE BINDING (BY SIMILARITY).
FT SITE 301 AA; 33186 MW; 28688E2ALC58EC9 CRC64;
SQ SEQUENCE 301 AA; 33186 MW; 28688E2ALC58EC9 CRC64;
Query Match 42.8%; Score 44.5; DB 1; Length 301;
Best Local Similarity 50.0%; Pred. No. 12;
Matches 10; Conservative 2; Mismatches 5; Indels 3; Gaps 1;
QY 3 RAVRPALSGDGR---VGSG 19
DB 133 QTIREKLGDFGLRVFGDG 152
RESULT 5
IF4A_CRYPV STANDARD; PRT; 405 AA.
AC O02494;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Eukaryotic initiation factor 4A (eif4A) (eif-4A).
GN EIF4-A.
OS Cryptosporidium parvum.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
OC Cryptosporidiidae; Cryptosporidium.
OX NCBI_TaxID=5807;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Moreduin;
RA Spano F., Putignani L., Criesanti A.;
RT "Cloning of the eif4A-A translation initiation factor gene of
RT Cryptosporidium parvum."
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: EIF4A is both a subunit of a high molecular weight
CC protein complex involved in cap recognition and is required as a
CC single polypeptide chain for mRNA binding to ribosome. It is an
CC ATP-dependent single stranded RNA-binding protein with a sequence-
CC independent unwinding activity (helicase).
CC -!- SUBUNIT: EIF4F is a multi-subunit complex, the composition of
CC which varies with external and internal environmental conditions.
CC It is composed of at least EIF4A, EIF4E and EIF4G (By similarity).
CC -!- SIMILARITY: Belongs to the DEAD box helicase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF001378; AAB58799.1; -.
DR EMBL; AF001211; AAB58726.1; -.
DR HSSP; Q58083; 1HV8.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR000629; DEAD_box.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.

DR PROSITE; PS00039; DEAD ATP HELICASE; 1.
KW Protein biosynthesis; Helicase; Initiation factor; ATP-binding;
KW DNA-binding; RNA-binding.
FT NP_BIND 75 82 ATP (BY SIMILARITY).
FT SITE 180 183 DEAD BOX.
SQ SEQUENCE 405 AA; 45933 MW; BD1E58048C6B79A CRC64;
Query Match 42.3%; Score 44; DB 1; Length 405;
Best Local Similarity 45.0%; Pred. No. 19;
Matches 9; Conservative 4; Mismatches 5; Indels 2; Gaps 1;
QY 2 QRAVRPALSGFD--GRVSG 19
DB 59 QRGKIPILDGYDTIGQAQSG 78
RESULT 6
LBPA_NEIMB STANDARD; PRT; 943 AA.
ID LBPA_NEIMB
AC Q06379; Q9JVK5;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Lactoferrin binding protein A precursor (Iron-regulated outer membrane
DE protein A).
GN LBPA OR IROA OR NMB1540.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ENCV / Serogroup B;
RX MEDLINE=94011384; PubMed=8406871;
RA Petteferson A., van der Ley P., Poolman J.T., Tommassen J.;
RT "Molecular characterization of the 98-kilodalton iron-regulated outer
RT membrane protein of Neisseria meningitidis."
RL Infect. Immun. 61:4724-4733 (1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Ciftone H., Clark E.B.,
RA Cotton M.D., Uterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Maignani V., Pizzo M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science 287:1809-1815 (2000).
CC -!- FUNCTION: UNKNOWN. MAY BE AN IRON-SIDEROPHORE RECEPTOR.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- INDUCTION: By iron starvation.
CC -!- SIMILARITY: Belongs to the tonB-dependent receptor family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X69214; CAA49148.1; -.
DR EMBL; AE002504; AAF41895.1; -.
DR FTR; G81070; G81070.
DR TIGR; NMB1540; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_dep_Rec; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.


```

FT CARBOHYD 882 882 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 895 895 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1213 1213 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1225 1225 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1267 1267 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1274 1274 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1355 AA; 149119 MW; 95E65A99E974CF63 CRC64;

Query Match 42.3%; Score 44; DB 1; Length 1355;
Best Local Similarity 62.5%; Pred. No. 64;
Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 3 RAVRPALSGDGRVGS 18
DB 256 RADRPVSTGVDCRVGS 271

RESULT 9
NTC4 MOUSE
ID NTC4 MOUSE STANDARD; PRT; 1964 AA.
AC P31635; O35442; O88314; O88316; Q62389; Q62390; Q9R1W9; Q9R1X0;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neurogenic locus notch homolog protein 4 precursor (Notch 4)
DE [Contains: Transforming protein Int-3].
DE NOTCH4 OR INT3 OR INT-3.
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92194507; PubMed=1312643;
RA Robbins J., Blondel B.J., Gallahan D., Callahan R.;
RT "Mouse mammary tumor gene int-3: a member of the notch gene family
transforming mammary epithelial cells.";
RL J. Virol. 66:2594-2599(1992).
RN [2]
RP REVISIONS, SEQUENCE FROM N.A.
RX MEDLINE=97294599; PubMed=9150355;
RA Gallahan D., Callahan R.;
RT "The mouse mammary tumor associated gene INT3 is a unique member of
the NOTCH gene family (NOTCH4).";
RL Oncogene 14:1883-1890(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX TISSUE=Lung, and Testis;
RC MEDLINE=96281668; PubMed=8681805;
RA Uyttendaele H., Marazzi G., Wu G., Yan Q., Sassoon D., Kitajewski J.;
RT "Notch4/int-3, a mammary proto-oncogene, is an endothelial
cell-specific mammalian Notch gene.";
RL Development 122:2251-2259(1996).
RN [4]
RP SEQUENCE FROM N.A.
RA Rowen L., Mahairas G., Qin S., Ahearn M.E., Dankers C., Lasky S.,
RA Loretz C., Schmidt S., Tipton S., Traicoff R., Zackrone K., Hood L.;
RT "Sequence of the mouse major histocompatibility locus class III
region.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 1436-1600 FROM N.A.
RX MEDLINE=99252212; PubMed=10233982;
RA Lee J.-S., Haruna T., Ishimoto A., Honjo T., Yanagawa S.-I.;
RT "Intracutaneous type A particle-mediated activation of the Notch4/int3
gene in a mouse mammary tumor: generation of truncated Notch4/int3
mRNAs by retroviral splicing events.";
RL J. Virol. 73:5166-5171(1999).
RN [6]
RP FUNCTION.
RX MEDLINE=21244657; PubMed=11344305;
RA Uyttendaele H., Ho J., Rossant J., Kitajewski J.;
RT "Vascular patterning defects associated with expression of activated

```

```

RT Notch4 in embryonic endothelium.";
RN Proc. Natl. Acad. Sci. U.S.A. 98:5643-5648(2001).
RL [7]
RP SEQUENCE OF 1463-1964, POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS
OF VAL-1463.
RX MEDLINE=21523956; PubMed=11518718;
RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
RT "Murine notch homologs (N1-4) undergo presenilin-dependent
proteolysis.";
RL J. Biol. Chem. 276:40268-40273(2001).
RN [8]
RP POST-TRANSLATIONAL PROCESSING.
RX MEDLINE=21374376; PubMed=11459941;
RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
RT "Conservation of the biochemical mechanisms of signal transduction
among mammalian Notch family members.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
CC -!- FUNCTION: Functions as a receptor for membrane-bound ligands
Jagged1, Jagged2 and Delta to regulate cell-fate determination.
Upon ligand activation through the released notch intracellular
domain (NICD) it forms a transcriptional activator complex with
RBP-J kappa and activates genes of the enhancer of split locus.
Affects the implementation of differentiation, proliferation and
apoptotic programs (by similarity). May regulate branching
morphogenesis in the developing vascular system.
CC -!- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-
terminal fragment N(EC) which are probably linked by disulfide
bonds.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein. Following
proteolytical processing NICD is translocated to the nucleus.
CC -!- TISSUE SPECIFICITY: Highly expressed in lung, moderately in heart
kidney, and at lower levels in the ovary and skeletal muscle. A
very low expression is seen in the brain, intestine, liver and
testis.
CC -!- DEVELOPMENTAL STAGE: Highly expressed in endothelial cells during
embryonic development from 9.0 dpc.
CC -!- PTM: Synthesized in the endoplasmic reticulum as an inactive form
which is proteolytically cleaved by a furin-like convertase in the
trans-Golgi network before it reaches the plasma membrane to yield
an active, ligand-accessible form. Cleavage results in a C-
terminal fragment N(TM) and a N-terminal fragment N(EC). Following
ligand binding, it is cleaved by TNF-alpha converting enzyme
(TACE) to yield a membrane-associated intermediate fragment called
notch extracellular truncation (NEXT). This fragment is then
cleaved by presenilin dependent gamma-secretase to release a
notch-derived peptide containing the intracellular domain (NICD)
from the membrane.
CC -!- PTM: Phosphorylated.
CC -!- DISEASE: Loss of the extracellular domain causes constitutive
activation of the Notch protein, which leads to hyperproliferation
of glandular epithelial tissues and development of mammary
carcinomas.
CC -!- SIMILARITY: Belongs to the NOTCH family.
CC -!- SIMILARITY: Contains 29 EGF-like domains.
CC -!- SIMILARITY: Contains 3 Lin/Notch repeats.
CC -!- SIMILARITY: Contains 5 ANK repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M80456; AAB38377.1; -
CC EMBL; U43691; AAC52630.1; -
CC EMBL; U43691; AAC52631.1; -
CC EMBL; AF030001; AAB82004.1; -
CC EMBL; AB016771; BAA32281.1; ALT_SEQ.
CC EMBL; AB016772; BAA32283.1; ALT_INIT.
CC EMBL; AB016773; BAA32284.1; ALT_INIT.
CC EMBL; AB016774; BAA32285.1; -

```


DR PIR; A38072; TWMT3.
DR PIR; T09059; T09059.
DR HSP; P08709; 1BF9.
DR MGD; MGI:107471; Notch4.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx_hydroxyl_s.
DR InterPro; IPR000742; EGF-2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR006209; EGF-like.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR008297; Notch_dom.
DR InterPro; IPR008000; Notch_dom.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00008; EGF; 27.
DR Pfam; PF00066; notch; 2.
DR PIRSF; PIRSF002379; Notch; 1.
DR PRINTS; PR00010; EGF_BLOOD.
DR PRINTS; PR00011; EGF_FLAMININ.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 6.
DR SMART; SM00179; EGF_Ca; 11.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS0287; ANK_REPEAT; 1..
DR PROSITE; PS0088; ANK_REPEAT; 5.
DR PROSITE; PS00010; ASX_HYDROXYL; 11.
DR PROSITE; PS00022; EGF_1; 28.
DR PROSITE; PS01186; EGF_2; 21.
DR PROSITE; PS00026; EGF_3; 27.
DR PROSITE; PS01187; EGF_Ca; 9.
KW Receptor; transcription regulation; Activator; Differentiation;
KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Glycoprotein; Signal; Phosphorylation; Proto-oncogene.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 1964 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 4.
FT CHAIN 1411 1964 TRANSFORMING PROTEIN INT-3.
FT CHAIN 1428 1964 NOTCH EXTRACELLULAR TRUNCATION.
FT CHAIN 1463 1964 NOTCH INTRACELLULAR DOMAIN.
FT DOMAIN 21 1443 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1444 1464 POTENTIAL.
FT DOMAIN 1465 1964 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 21 60 EGF-LIKE 1.
FT DOMAIN 61 112 EGF-LIKE 2.
FT DOMAIN 115 152 EGF-LIKE 3.
FT DOMAIN 153 189 EGF-LIKE 4.
FT DOMAIN 191 229 EGF-LIKE 5.
FT DOMAIN 231 271 EGF-LIKE 6.
FT DOMAIN 273 309 EGF-LIKE 7.
FT DOMAIN 311 350 EGF-LIKE 8.
FT DOMAIN 352 388 EGF-LIKE 9.
FT DOMAIN 389 427 EGF-LIKE 10.
FT DOMAIN 429 470 EGF-LIKE 11.
FT DOMAIN 472 508 EGF-LIKE 12.
FT DOMAIN 510 546 EGF-LIKE 13.
FT DOMAIN 548 584 EGF-LIKE 14.
FT DOMAIN 586 622 EGF-LIKE 15.
FT DOMAIN 623 656 EGF-LIKE 16.
FT DOMAIN 658 686 EGF-LIKE 17.
FT DOMAIN 688 724 EGF-LIKE 18.
FT DOMAIN 726 762 EGF-LIKE 19.
FT DOMAIN 764 800 EGF-LIKE 20.
FT DOMAIN 803 839 EGF-LIKE 21.
FT DOMAIN 841 877 EGF-LIKE 22.
FT DOMAIN 878 924 EGF-LIKE 23.
FT DOMAIN 926 962 EGF-LIKE 24.
FT DOMAIN 964 1000 EGF-LIKE 25.
FT DOMAIN 1002 1040 EGF-LIKE 26.
FT DOMAIN 1042 1081 EGF-LIKE 27.
FT DOMAIN 1083 1122 EGF-LIKE 28.
FT DOMAIN 1126 1167 EGF-LIKE 29.
FT REPEAT 1168 1208 LIN/NOTCH 1.
FT REPEAT 1209 1242 LIN/NOTCH 2.
FT REPEAT 1243 1282 LIN/NOTCH 3.

Query Match 42.3%; Score 44; DB 1; Length 1964;
Best Local Similarity 52.9%; Pred. No. 93;
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;
QY 3 RAVRPALSGFDCRGVSG 19
DB 1165 RCORPGACGCGRGD 1181
RESULT 10
YHGI_ECOLI
ID YHGI_ECOLI STANDARD; PRT; 191 AA.
AC P46847;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Protein yhgI.
GN YHGI OR B3414 OR C4191 OR Z4769 OR ECS4256 OR SF3437 OR S4328.
OS Escherichia coli,
OS Escherichia coli O6,
OS Escherichia coli O157:H7, and
OS Shigella flexneri
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 217992, 83334, 623;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=K12 / MGI655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz F.R.;
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Godbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RL Nature 409:529-533 (2001).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Tada T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22 (2001).
RN [5]
RP SEQUENCE FROM N.A.
RC SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;

```

RX MEDLINE=22272406; PubMed=12384530;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
RT through comparison with genomes of Escherichia coli K12 and O157."
RL Nucleic Acids Res. 30:4432-4441(2002).
[6]
SEQUENCE FROM N.A.
RC SPECIES=S.flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
RT flexneri serotype 2a strain 2457T."
RL Infect. Immun. 71:2775-2786(2003).
CC -!- SIMILARITY: STRONG, TO H. INFLUENZAE HI0433 AND B. APHIDICOLA
CC (SUBSP. ACYRTHOSIPHON PISUM) BU544.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; U18997; AAA58212.1; -
DR EMBL; AE000417; AAC76439.1; -
DR EMBL; AB016768; AAN82629.1; -
DR EMBL; AE005563; AAG58515.1; -
DR EMBL; AP002565; BAB37679.1; -
DR EMBL; AE015351; AAN44898.1; -
DR EMBL; AE016992; AAP19283.1; -
DR PIR; A65137; A65137.
DR PIR; G86006; G86006.
DR PIR; H91160; H91160.
DR SWISS-2DPAGE; P46847; COI1.
DR EcoGene; EG12935; YhgI.
DR InterPro; IPR000361; HesB_yadR_yfhp.
DR InterPro; IPR001075; NifU_C.
DR Pfam; PF01521; HesB-like; 1.
DR Pfam; PF01106; NifU-like; 1.
DR ProDom; PD002830; NifU_C; 1.
DR Complete proteome.
KW SEQUENCE 191 AA; 20998 MW; 06874546ADA5A971 CRC64;
SQ
Query Match 41.8%; Score 43.5; DB 1; Length 191;
Best Local Similarity 31.4%; Pred. No. 11;
Matches 11; Conservative 2; Mismatches 7; Indels 15; Gaps 1;
QY 1 LQRAVRPALSGPGRV-----GSGC 20
DB 115 LQSQINPQLAGHGGRVSLMEITDGYAILQFGGCG 149
RESULT 11
REQO_BPAPS STANDARD; PRT; 148 AA.
AC Q9TJ03;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Probable antitermination protein Q (P5).
GN 5
OS Bacteriophage APSE-1.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae.
OX NCBI_TaxID=106199;
RN [1]
```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=99420383; PubMed=10489345;
RA van der Wilk F., Dullemeans A.M., Verbeek M., van den Heuvel J.F.J.M.;
RT "Isolation and characterization of APSE-1, a bacteriophage infecting
RT the secondary endosymbiont of acyrtosiphon pisum."
RL Virology 262:104-113(1999).
CC -!- FUNCTION: Positively regulates expression of some phage genes.
CC Bacterial host RNA polymerase modified by antitermination proteins
CC transcribes through termination sites that otherwise prevent
CC expression of the regulated genes (By similarity).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AF157835; AAF03999.1; -
KW Transcription regulation; Transcription antitermination; DNA-binding.
SQ SEQUENCE 148 AA; 16889 MW; 93F1D758A8505170 CRC64;
Query Match 40.4%; Score 42; DB 1; Length 148;
Best Local Similarity 40.0%; Pred. No. 14;
Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;
QY 1 LQRAVRPALSGPGRVSGSC 20
DB 38 LRPSMRPSCSDGRIIDNC 57
RESULT 12
YPS7_MYCTU STANDARD; PRT; 224 AA.
AC Q50741;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein RV2557/MT2634/MB2587.
GN RV2557 OR MT2634 OR MTCY9C4.11C OR MB2587.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaiia F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean L.A., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
[2]
SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL J. Bacteriol. 184:5479-5490(2002).
```



```

RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396508; PubMed=11481431;
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
RA Vorhoefer F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,
RA Golding B., Puehler A.,
RT the complete sequence of the 1.693-kb pSymb megaplasmid from the N2-
RT fixing endosymbiont Sinorhizobium meliloti.;
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
CC -!- CATALYTIC ACTIVITY: N-(L-arginino)succinate = fumarate + L-
CC arginine.
CC -!- PATHWAY: Arginine biosynthesis; eighth (last) step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- SIMILARITY: Belongs to the lyase 1 family. Argininosuccinate lyase
CC subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL603644; CAC49073.1; -.
CC PIR; A95926; A95926.
CC HAMAP; MF 00006; -.
CC InterPro; IPR009049; argH.
CC InterPro; IPR000362; Fumarate lyase.
CC InterPro; IPR008948; L-Aspartase-like.
CC Pfam; PF00206; lyase_1; 1.
CC PRINTS; PR00149; FUMARATELYASE.
CC PROSITE; PS00838; argH; 1.
CC PROSITE; PS00163; FUMARATE LYASES; 1.
KW Arginine biosynthesis; Lyase; Plasmid; Complete proteome.
SQ SEQUENCE 488 AA; 53037 MW; 376988B050516832 CRC64;

Query Match 40.4%; Score 42; DB 1; Length 488;
Best Local Similarity 50.0%; Pred. No. 47;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 4 AVRPALSGDGRVGS 19
DB 61 AIRAALLEGTEADVG 76

RESULT 17
METK_TREPA STANDARD; PRT; 396 AA.
AC O83772;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE S-adenosylmethionine synthetase (EC 2.5.1.6) (Methionine
DE adenosyltransferase) (AdoMet synthetase) (MAT).
GN METK OR TP0794.
OS Treponema pallidum.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Nichols;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artiaich P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete.";
RT Science 281:375-388(1998).

```

```

CC -!- FUNCTION: Catalyzes the formation of S-adenosylmethionine from
CC methionine and ATP. The overall synthetic reaction is composed of
CC two sequential steps, AdoMet formation and the subsequent
CC triphosphosphate hydrolysis which occurs prior to release of
CC AdoMet from the enzyme (By similarity).
CC -!- CATALYTIC ACTIVITY: ATP + L-methionine + H(2)O = phosphate +
CC diphosphate + S-adenosyl-L-methionine.
CC -!- COFACTOR: Binds 2 divalent ions, such as magnesium or cobalt, and
CC 1 potassium ion per subunit (By similarity).
CC -!- PATHWAY: Activated methyl cycle.
CC -!- SUBUNIT: Homotetramer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the AdoMet synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; A5001250; AAC65758.1; -.
CC PIR; A71281; A71281.
CC HSSP; P04384; 1FUG.
CC TIGR; TP0794; -.
CC HAMAP; MF 00086; -.
CC InterPro; IPR002133; S-AdoMet synt.
CC Pfam; PF00438; S-AdoMet synt_1.
CC Pfam; PF02772; S-AdoMet syntD2; 1.
CC Pfam; PF02773; S-AdoMet syntD3; 1.
CC TIGRFAMs; TIGR01034; metK; 1.
CC PROSITE; PS00376; ADO MET SYNTHETASE 1; 1.
CC PROSITE; PS00377; ADO MET SYNTHETASE 2; 1.
KW Transferase; One-carbon metabolism; ATP-binding; Magnesium; Potassium;
KW Metal-binding; Complete proteome.
FT NP_BIND 264 271 ATP (POTENTIAL).
FT METAL 16 16 MAGNESIUM (BY SIMILARITY).
FT METAL 42 42 POTASSIUM (BY SIMILARITY).
FT METAL 268 268 POTASSIUM (BY SIMILARITY).
FT METAL 276 276 MAGNESIUM (BY SIMILARITY).
SQ SEQUENCE 396 AA; 42990 MW; E532DD39B0BBE711 CRC64;

Query Match 39.9%; Score 41.5; DB 1; Length 396;
Best Local Similarity 91.7%; Pred. No. 46;
Matches 11; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 LQAVRVP-ALSG 11
DB 381 LQAVRVPFALSG 392

RESULT 18
VP19_HSV14 STANDARD; PRT; 465 AA.
AC P17586; P10222;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Capsid assembly and DNA maturation protein (Virion protein UL38)
DE (Capsid protein VP19C).
GN UL38.
OS Herpes simplex virus (type 1 / strain A44).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Simplexvirus.
OX NCBI_TaxID=10300;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89199774; PubMed=2539510;
RA Pertuiset B., Boccara M., Cebrian J., Berthelot N., Chousterman S.,
RA Pavion-Duttilleul F., Sisman J., Sheldrick P.;
RT "Physical mapping and nucleotide sequence of a herpes simplex virus
RT type 1 gene required for capsid assembly.";

```



```

RT genomic discontinuities and transcript suggest that it is a
RL pararetrovirus."
CC Nucleic Acids Res. 18:5505-5513(1990).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; X52938; CAA37109.1; -
DR PIR; S11478; S11478.
KW Hypothetical protein.
SQ SEQUENCE 135 AA; 14786 MW; 02216F2C91D4AF21 CRC64;

Query Match 39.4%; Score 41; DB 1; Length 135;
Best Local Similarity 80.0%; Pred. No. 19;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 QRAVPAISG 11
DB 103 QRAVPAISG 112

RESULT 22
DAPB METHTH STANDARD; PRT; 273 AA.
AC Q26891;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dihydrodipicolinate reductase [EC 1.3.1.26] (DHPR).
GN DAPB OR MTH800.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteriia; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Delta H;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiواني N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT delcAH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
CC -!- CATALYTIC ACTIVITY: 2,3,4,5-tetrahydrodipicolinate + NAD(P) (+) =
CC 2,3-dihydrodipicolinate + NAD(P)H.
CC -!- PATHWAY: Biosynthesis of diaminopimelate and lysine from aspartate
CC semialdehyde; second step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the dihydrodipicolinate reductase family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AE000858; BAB85300.1; -
DR PIR; H69206; H69206.
DR HSSP; P04036; 1DRW.
DR HAMAP; MF_00102; -; 1.
DR InterPro; IPR000846; DapB.

```

```

DR Pfam; PF05173; DapB_C; 1.
DR Pfam; PF01113; DapB_N; 1.
DR ProDom; PD004105; DapB; 1.
DR TIGRfams; TIGR00036; dapB; 1.
DR PROSITE; PS01298; DAPB; 1.
KW Diaminopimelate biosynthesis; Lysine biosynthesis; Oxidoreductase;
KW NADP; Complete proteome.
SQ SEQUENCE 273 AA; 29045 MW; 5242C2A5D7B59B15 CRC64;

Query Match 39.4%; Score 41; DB 1; Length 273;
Best Local Similarity 53.3%; Pred. No. 38;
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 5 VRPALSGFDGRVSG 19
DB 2 IRVAVTGACGRMGSG 16

RESULT 23
ROC_XENLA STANDARD; PRT; 282 AA.
AC P19600;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Heterogeneous nuclear ribonucleoprotein C (hnRNP core protein C).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=83355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89071757; PubMed=2904678;
RA Preugschat F., Wold B.;
RT "Isolation and characterization of a Xenopus laevis C protein cDNA:
RT structure and expression of a heterogeneous nuclear ribonucleoprotein
RT core protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:9669-9673(1988).
CC -!- FUNCTION: May play a role in nucleosome assembly by neutralizing
CC basic proteins such as A and B core hnRNPs.
CC -!- SUBCELLULAR LOCATION: Nuclear; component of ribonucleosomes.
CC -!- PTM: Phosphorylated (Probable).
CC -!- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; J03831; AAA60937.1; -
DR PIR; A31765; A31765.
DR InterPro; IPR005054; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS50102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW Nuclear protein; RNA-Binding; Ribonucleoprotein; Phosphorylation.
FT DOMAIN 17 88
FT NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 141 147
FT ASP/GLU-RICH (ACIDIC).
FT DOMAIN 178 282
FT PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 240 240
FT PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 256 256
FT MOD_RES 267 267
FT SEQUENCE 282 AA; 30950 MW; 7373FA46F8C85413 CRC64;

Query Match 39.4%; Score 41; DB 1; Length 282;
Best Local Similarity 44.4%; Pred. No. 39;
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

```



```

OY 2 QRAVRPALSGFDGRVCSG 19
DB 61 ERTARTAVAGEDGRMTAG 78

RESULT 24
SPB2_HUMAN STANDARD; PRT; 953 AA.
AC Q14151; Q8TB13;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Scaffold attachment factor B2.
GN SAFB2 OR KIAA0138.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127530; PubMed=8590280;
RA Nagase T., Seki N., Tanaka A., Ishikawa K.-I., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. IV.
RT The coding sequences of 40 new genes (KIAA0121-KIAA0160) deduced by
RT analysis of cDNA clones from human cell line KG-1."
RL DNA Res. 2:167-174(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,
RA Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stilwagen S.,
RA Phan H., Velasco N., Ganes J., Danganan L., Poundstone P.,
RA Christensen M., Georgescu A., Avila J., Liu S., Attix C., Andreise T.,
RA Trankheim M., Amico-Keller G., Coefield J., Duarte S., Lucas S.,
RA Bruce R., Thomas P., Quan G., Krommiller B., Arellano A.,
RA Montgomery M., Ow D., Nolan M., Trong S., Kobayashi A., Olsen A.O.,
RA Carrano A.V.;
RA Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RL [3]
RP SEQUENCE OF 1-528 FROM N.A.
RX TISSUE=Pancreas;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zebberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
RA Schnurch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=22651121; PubMed=12660241;
RA Townson S.M., Dobrzycka K.M., Lee A.V., Air M., Deng W., Kang K.,
RA Jiang S., Kioha N., Michaelis K., Oesterreich S.;
RT "SAFB2, a new scaffold attachment factor homolog and estrogen receptor
RT corepressor."
RL J. Biol. Chem. 278:20059-20068(2003).
CC -!- FUNCTION: Binds to scaffold/matrix attachment region (S/MAR) DNA.
CC Can function as an estrogen receptor co-repressor and can also
CC inhibit cell proliferation.
CC -!- SUBUNIT: Interacts with SAFB/SAFB1 and SCAM1.

```

-!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.
 -!- TISSUE SPECIFICITY: Expressed at high levels in the CNS and at low levels in the liver. Expressed in a wide number of breast cancer cell lines.
 -!- PTM: Phosphorylated (By similarity).
 -!- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.
 -!- SIMILARITY: Contains 1 SAP domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; D50928; BAA09487.1; -;
 EMBL; AC004611; AAC14666.1; -;
 EMBL; BC025279; AAH25279.1; -;
 Genew; HGNC:21605; SAFB2.
 MIM; 608066; -;
 InterPro; IPR000504; RNA_rec_mot.
 InterPro; IPR003034; SAP.
 Pfam; PF00076; rrm; 1.
 Pfam; PF02037; SAP; 1.
 SMART; SM00360; RRM; 1.
 SMART; SM00513; SAP; 1.
 PROSITE; PS0102; RRM; 1.
 PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
 PROSITE; PS00800; SAP; 1.
 Transcription regulation; Repressor; Nuclear protein; DNA-binding;
 RNA-binding; Phosphorylation.

FT DOMAIN 30 64 SAP.
 FT DOMAIN 407 485 RNA-BINDING (RRM).
 FT DOMAIN 600 953 INTERACTS WITH SAFB1.
 FT DOMAIN 713 730 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 482 545 LYS-RICH.
 FT DOMAIN 619 724 GLU-RICH.
 FT DOMAIN 621 788 ARG-RICH.
 FT DOMAIN 792 926 GLY-RICH.
 FT CONFLICT 528 528 K -> M (IN REF. 2).
 SQ SEQUENCE 953 AA; 107473 MW; 084343934F8B3196 CRC64;

Query Match 39.4%; Score 41; DB 1; Length 953;
 Best Local Similarity 38.9%; Pred. No. 1.3e+02;
 Matches 7; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 2 QRAVRPALSGFDGRVCSG 19
 DB 387 EKDKPIIKDKGRVCSG 404
 : : : : :
 : : : : :

RESULT 25
 ARVC_MOUSE STANDARD; PRT; 969 AA.
 AC P98203;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Armadillo repeat protein deleted in velo-cardio-facial syndrome
 DE homolog (Fragment).
 GN ARVCF.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20512094; PubMed=11058098;
 RA Kaufmann U., Zupping C., Waibler Z., Rudiger M., Urbich C.,
 RA Martin B., Jockusch B.M., Eppenberger H., Starzinski-Powitz A.;
 RT "The armadillo repeat region targets ARVCF to cadherin-based cellular junctions."

```

J. Cell Sci. 113:4121-4135(2000).
CC - FUNCTION: INVOLVED IN PROTEIN-PROTEIN INTERACTIONS AT ADHERENS
CC - JUNCTIONS (BY SIMILARITY).
CC - SIMILARITY: Belongs to the beta-catenin family.
CC - SIMILARITY: Contains 10 ARM repeats.
CC - SIMILARITY: Contains 10 ARM repeats.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL; AJ243418; -; NOT_ANNOTATED_CDS.
CC HSSP; F35222; IG3J.
CC MGD; MGI:109620; Arvcf.
CC GO; GO:0005737; C:cytoplasm; IDA.
CC GO; GO:0005634; C:nucleus; IDA.
CC GO; GO:0005886; C:plasma membrane; IDA.
CC GO; GO:0005515; F:protein binding; IPI.
CC InterPro; IPR008938; ARM.
CC Pfam; PF00514; Armadillo_seg. 4.
CC SMART; SM00185; ARM; 5.
CC PROSITE; PS0176; ARM_REPEAT; 3.
KW Cell adhesion; Cytoskeleton; Structural protein; Repeat; Coiled coil.
FT NON_TER 1 1
FT DOMAIN <1 39 COILED COIL (POTENTIAL).
FT REPEAT 328 376 ARM 1.
FT REPEAT 377 418 ARM 2.
FT REPEAT 419 462 ARM 3.
FT REPEAT 463 503 ARM 4.
FT REPEAT 521 560 ARM 5.
FT REPEAT 570 616 ARM 6.
FT REPEAT 634 674 ARM 7.
FT REPEAT 675 720 ARM 8.
FT REPEAT 721 769 ARM 9.
FT REPEAT 770 814 ARM 10.
SQ SEQUENCE 969 AA; 105378 MW; 91130069E484C5B5 CRC64;

Query Match 38.9%; Score 40.5; DB 1; Length 969;
Best Local Similarity 52.6%; Pred. No. 1.6e+02;
Matches 10; Conservative 3; Mismatches 5; Indels 1; Gaps 1;

QY 1 LQRAVRPALSGFDGRVGS 19
Db 34 LERAQQPGMSS-GGMVGS 51

RESULT 26
HXAB_MOUSE
ID HXAB_MOUSE STANDARD; PRT; 313 AA.
AC P31311;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Homeobox protein Hox-A11 (Hox-1.9).
GN HOXA11 OR HOXA-11 OR HOX-1.9.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=95309120; PubMed=7789268;
RA Haieh-Li H.M., Witte D.P., Weinstein M., Branford W., Li H., Small K.,
RA Potter S.S.;
RT "Hoxa 11 structure, extensive antisense transcription, and function
RT in male and female fertility."
RL Development 121:1373-1385(1995).
RN [2]
RP SEQUENCE OF 241-300 FROM N.A.

J. Cell Sci. 113:4121-4135(2000).
CC - FUNCTION: INVOLVED IN PROTEIN-PROTEIN INTERACTIONS AT ADHERENS
CC - JUNCTIONS (BY SIMILARITY).
CC - SIMILARITY: Belongs to the beta-catenin family.
CC - SIMILARITY: Contains 10 ARM repeats.
CC - SIMILARITY: Contains 10 ARM repeats.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL; U20371; AAA85711.1; -.
CC EMBL; U20370; AAA85710.1; -.
CC F01R; G37290; G37290.
CC HSSP; PL4653; 1B72.
CC TRANSFAC; T01717; -.
CC MGD; MGI:96172; Hoxa11.
CC GO; GO:0001501; P:skeletal development; IMP.
CC InterPro; IPR001356; Homeobox.
CC Pfam; PF00046; homeobox; 1.
CC PRINTS; PR00024; HOMEBOX.
CC ProDom; PD000010; Homeobox; 1.
CC SMART; SM00389; HOX; 1.
CC PROSITE; PS00027; HOMEBOX_1; 1.
CC PROSITE; PS50071; HOMEBOX_2; 1.
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;
KW Transcription regulation.
FT DOMAIN 172 175 POLY-ALA.
FT DOMAIN 178 185 POLY-ALA.
FT DOMAIN 196 201 POLY-GLY.
FT DOMAIN 212 215 POLY-ARG.
FT DOMAIN 218 221 POLY-SER.
FT DNA_BIND 241 300 HOMEBOX.
SQ SEQUENCE 313 AA; 34484 MW; 07BE80F8E2225B3E CRC64;

Query Match 38.5%; Score 40; DB 1; Length 313;
Best Local Similarity 57.1%; Pred. No. 62;
Matches 8; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 7 PALSGFDGRVGS 20
Db 189 PATSSSDGGGGGC 202

RESULT 27
SPS2_HUMAN
ID SPS2_HUMAN STANDARD; PRT; 448 AA.
AC Q99611; Q9BUQ2;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Selenide, water dikinase 2 (SC 2.7.9.3) (Selenophosphate synthetase 2)
DE (Selenium donor protein 2).
GN SEPHS2 OR SPS2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=96017645; PubMed=7588067;
RA Guimaraes M.J., Bazan J.F., Zlotnik A., Wiles M.V., Grimaldi J.C.,
RA Lee F., McClanahan T.;
RT "A new approach to the study of haematopoietic development in the
RT yolk sac and embryoid bodies."

```

Development 121:3335-3346(1995).

[2]
 RL SEQUENCE FROM N.A.
 RN MEDLINE=97140286; PubMed=8986768;
 RX Guinard M.J., Peterson D., Vicari A., Cocks B.G., Copeland N.G.,
 RA Gilbert D.J., Jenkins N.A., Ferrick D.A., Kastelein R., Bazan J.F.,
 RA Zlotnik A.;
 RT "Identification of a novel seld homolog from eukaryotes, bacteria,
 RT and archaea: is there an autoregulatory mechanism in selenocysteine
 RT metabolism?";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:15086-15091(1996).

[3]
 RN SEQUENCE FROM N.A.
 RP TISSUE=Muscle, and Skin;
 RC MEDLINE=22388257; PubMed=12477932;
 RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zebberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Tothiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S., Krzywicki M.I., Skalska U.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

[4]
 CC -!- FUNCTION: Synthesizes selenophosphate from selenide and ATP.
 CC -!- CATALYTIC ACTIVITY: ATP + selenide + H(2)O = AMP + selenophosphate
 CC + phosphate.
 CC -!- COFACTOR: Selenocysteine. The active-site selenocysteine is
 CC encoded by the opal codon, UGA.
 CC -!- SIMILARITY: Belongs to the selenophosphate synthetase 1 family.
 CC Class I subfamily.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U43286; AAC50958.2; -;
 DR EMBL; BC002381; RA002381.3; -;
 DR EMBL; BC016643; AAH16643.1; -;
 CC MIM; 606218; -;
 DR GO; GO:0004756; F:selenide, water dikinase activity; NAS.
 DR GO; GO:0016260; P:selenocysteine biosynthesis; NAS.
 DR InterPro; IPR000728; AIR synth.
 DR InterPro; IPR004536; Seld.
 DR Pfam; PF00586; AIRS; 1.
 DR Pfam; PF02769; AIRS; 1.
 DR TIGRFAMs; TIGR00476; seld; 1.
 KW Transferase; Selenium; Selenocysteine; ATP-binding.
 FT ACT_SITE 60 60
 FT ACT_SITE 60 60
 FT SITE 63 63
 FT SITE 63 63
 FT NP_BIND 319 325
 FT ATP (POTENTIAL).
 FT SIMILARITY.
 FT SEQUENCE 448 AA; 47258 MW; 343A58CD9F842B99 CRC64;
 SQ

Query Match 38.5%; Score 40; DB 1; Length 448;
 Best Local Similarity 58.3%; Pred. No. 90;
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 9 LSGFDRVSGC 20
 :|||
 Db 51 LTGFSGMGCGC 62

RESULT 28
 ALA9 ARATH
 ID ALA9 ARATH STANDARD; PRT; 1200 AA.
 AC Q9SX33;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Potential phospholipid-transporting ATPase 9 (EC 3.6.3.1)
 DE (Aminophospholipid flippase 9).
 GN ALA9 OR ATIG68710 OR F24J5.6.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
 RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
 RA Dunn P., Egu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.L., Krenetskaia I., Kurtz D.B., Kwan A., Lam B.,
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
 RA Miltischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 RT thaliana";
 RL Nature 408:816-820(2000).
 CC -!- FUNCTION: Involved in transport of phospholipids (Potential).
 CC -!- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- SIMILARITY: Belongs to the cation transport ATPases family (P-type
 CC ATPases). Subfamily IV.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AC008075; AAD49973.1; -;
 DR FIR; F96711; F96711.
 DR InterPro; IPR001757; ATPase E1-E2.
 DR InterPro; IPR006539; Flippase.
 DR InterPro; IPR005834; Hydrolase.
 DR Pfam; PF00702; Hydrolase; 1.
 DR PRINTS; PR00119; CATATPASE.
 DR TIGRFAMs; TIGR01652; ATPase-Plipid; 1.
 DR TIGRFAMs; TIGR01494; ATPase P-type; 6.
 DR PROSITE; PS00154; ATPase E1 E2; 1.
 KW Hydrolase; Transmembrane; Phosphorylation; ATP-binding; Metal-binding;
 FT DOMAIN 1 75
 FT TRANSMEM 76 97
 FT DOMAIN 98 101
 FT EXTRACELLULAR (POTENTIAL).

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2004, 06:49:47 ; Search time 33.3333 Seconds
(without alignments)
189.311 Million cell updates/sec

Title: ' 09549186-7

Perfect score: 104

Sequence: 1 LQRAVRPALSGFGRVGSQC 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

SPTREMBL25:

1: sp_archaea:

2: sp_bacteria:

3: sp_fungi:

4: sp_human:

5: sp_invertebrate:

6: sp_mammal:

7: sp_mhc:

8: sp_organelle:

9: sp_phase:

10: sp_plant:

11: sp_rodent:

12: sp_virus:

13: sp_vertebrate:

14: sp_unclassified:

15: sp_rvirus:

16: sp_bacteriaph:

17: sp_archaeap:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	53	51.0	404	2	Q93AD0
2	50	48.1	2879	5	Q9U0Z1
3	49	47.1	707	5	Q8MT57
4	49	47.1	818	5	Q9GYW8
5	49	47.1	818	5	Q9W278
6	48	46.2	150	12	Q91EQ8
7	48	46.2	150	12	Q91EQ8
8	48	46.2	376	16	Q9RYP2
9	48	46.2	438	2	Q845V3
10	47.5	45.7	191	16	Q98NT3
11	47.5	45.7	283	16	Q98AE8
12	47.5	45.7	283	16	Q982P9
13	47	45.2	240	16	Q83A70
14	46	44.2	166	10	Q8W5C0
15	45.5	43.8	1482	16	Q8EHA4
16	45	43.3	287	16	Q83611

17	45	43.3	299	16	Q98KT8
18	45	43.3	373	10	Q7XXP6
19	45	43.3	379	10	Q7XM00
20	45	43.3	398	5	Q9U937
21	45	43.3	398	5	Q8IKF0
22	45	43.3	405	10	Q9AYM0
23	45	43.3	405	10	Q7X976
24	45	43.3	453	10	Q9LKR6
25	45	43.3	484	10	Q8LQF2
26	45	43.3	484	16	Q8XDM4
27	45	43.3	484	16	Q8FCB4
28	45	43.3	484	16	Q83J34
29	45	43.3	499	5	Q86B00
30	45	43.3	1553	16	Q7UWH2
31	44.5	42.8	596	16	Q8PHF2
32	44	42.3	136	16	Q98J09
33	44	42.3	136	16	Q92ZC0
34	44	42.3	477	17	Q9HRL0
35	44	42.3	740	16	Q9RJD7
36	44	42.3	940	2	Q51187
37	44	42.3	943	2	Q87343
38	44	42.3	943	2	Q50952
39	44	42.3	1746	16	Q9HXR4
40	43.5	41.8	191	16	Q8ZLI7
41	43.5	41.8	191	16	Q8ZJ10
42	43.5	41.8	191	16	Q8Z223
43	43	41.3	217	16	Q7U831
44	43	41.3	298	16	Q82KA4
45	43	41.3	408	16	Q7VGM3
46	43	41.3	447	10	Q9PR32
47	43	41.3	606	10	Q8LMR1
48	43	41.3	688	5	Q879L6
49	43	41.3	1227	5	Q9W0M7
50	43	41.3	1747	5	Q26640
51	43	41.3	2809	4	Q96JP8
52	43	41.3	2809	4	Q86SJ5
53	42	40.4	104	2	Q50013
54	42	40.4	135	16	Q8UAA7
55	42	40.4	143	16	Q8UFP7
56	42	40.4	173	16	Q8XR81
57	42	40.4	174	5	Q9VKH4
58	42	40.4	176	17	Q9YDB1
59	42	40.4	246	17	Q8ZU06
60	42	40.4	272	5	Q9VE12
61	42	40.4	304	16	Q89HH5
62	42	40.4	330	16	Q9A6A7
63	42	40.4	333	16	Q7V4X2
64	42	40.4	404	4	Q12984
65	42	40.4	465	12	Q9QDK9
66	42	40.4	466	10	Q9SNC9
67	42	40.4	467	2	Q8GBZ5
68	42	40.4	472	16	Q8D9F3
69	42	40.4	478	16	Q87PA5
70	42	40.4	485	4	Q96GW2
71	42	40.4	497	16	Q93JB9
72	42	40.4	575	17	Q28180
73	42	40.4	606	10	Q7XUD3
74	42	40.4	616	16	Q9A3D0
75	42	40.4	689	17	Q8Q0K8
76	42	40.4	710	17	Q8TKV5
77	42	40.4	799	16	Q8GSU6
78	42	40.4	847	16	Q8DHS8
79	42	40.4	872	5	Q9NL45
80	42	40.4	898	2	Q85081
81	42	40.4	1037	16	Q7WCS7
82	42	40.4	1037	16	Q7VZR4
83	42	40.4	1114	16	Q8PGR8
84	42	40.4	1328	2	Q9LAX0
85	42	40.4	1396	3	Q8WZY9
86	42	40.4	1399	16	Q826G9
87	42	40.4	1421	9	Q9FZU3
88	42	40.4	2190	16	Q8PGS0
89	42	40.4	2351	16	Q8PCQ5

Q98kt8 rhizobium l	
Q7xsp6 oryza sativ	
Q7xm00 oryza sativ	
Q9u937 plasmodium	
Q8ikf0 plasmodium	
Q9aym0 oryza sativ	
Q7x976 oryza sativ	
Q9lkr6 arabidopsis	
Q8lqf2 oryza sativ	
Q8xdm4 escherichia	
Q8fce4 escherichia	
Q83j34 shigella fl	
Q86b00 dictyosteli	
Q7uwh2 rhodospirell	
Q8php2 xanthomonas	
Q98j09 rhizobium l	
Q92zc0 rhizobium m	
Q9rl0 halobacteri	
Q9rjd7 streptomyce	
Q51187 neisseria m	
Q87343 neisseria m	
Q50952 neisseria g	
Q9hxr4 pseudomonas	
Q8zli7 salmonella	
Q8zj10 yersinia pe	
Q8z223 salmonella	
Q7u831 synechococc	
Q82ka4 streptomyce	
Q7vgm3 helicobacte	
Q9f32 lycopersico	
Q8lmr1 oryza sativ	
Q879l6 drosophila	
Q9w0m7 drosophila	
Q26640 strongyloce	
Q96jp8 homo sapien	
Q86aj5 homo sapien	
Q50013 mycobacteri	
Q8uaa7 agrobacteri	
Q8ufp7 agrobacteri	
Q8xr81 ralstonia s	
Q9vkh4 drosophila	
Q9ydb1 aeropyrum p	
Q8azu6 pyrobaculum	
Q9ve12 drosophila	
Q89h5 bradyrhizob	
Q9a6a7 caulobacter	
Q7v4x2 prochloroco	
Q12984 homo sapien	
Q9qdk9 pterostyllis	
Q9anc9 arabidopsis	
Q8gbz5 nostoc punc	
Q8d9f3 vibrio vuln	
Q87pa5 vibrio para	
Q96gw2 homo sapien	
Q93jb9 streptomyce	
Q28180 archaeglob	
Q7xud3 oryza sativ	
Q9a3d0 caulobacter	
Q8q0k8 methanosarc	
Q8skv5 methanosarc	
Q8gsu6 bifidobacte	
Q8dh58 synechococc	
Q9nl45 ciona intes	
Q85081 moraxella c	
Q7wcs7 bordetella	
Q7vzr4 bordetella	
Q8pgr8 xanthomonas	
Q9lax0 xanthomonas	
Q8wzy9 neurospora	
Q826g9 streptomyce	
Q9fzuz3 neisseria m	
Q8pgs0 xanthomonas	
Q8pcq5 xanthomonas	

90 41.5 39.9 73 16 Q8EJX7
91 41.5 39.9 101 17 Q9YB56
92 41.5 39.9 283 10 Q22841
93 41.5 39.9 298 11 Q9DC93
94 41.5 39.9 298 11 Q9CQX9
95 41.5 39.9 327 11 Q9D7G7
96 41.5 39.9 343 11 Q9CZ42
97 41.5 39.9 482 16 Q9JWK8
98 41.5 39.9 483 16 Q9JXD2
99 41.5 39.9 621 2 Q59301
100 41.5 39.9 986 16 Q7UBE7

ALIGNMENTS

RESULT 1

Q93AD0 PRELIMINARY; PRT; 404 AA.
AC Q93AD0
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Putative penicillin binding protein Pbp2.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RA Fernandez A., Harwood C.S.;
RT "VanK, a vanillate transporter from Pseudomonas putida PRS2000.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF425229; AAL27557.1; --
DR GO; GO:0008658; F:penicillin binding; IEA.
DR InterPro; IPR005311; PBP dimer.
DR Pfam; PF03717; PBP dimer; 1.
DR
SQ SEQUENCE 404 AA; 44749 MW; 458F7770299FA78C CRC64;

Query Match 51.0%; Score 53; DB 2; Length 404;
Best Local Similarity 58.8%; Pred. No. 4.9;
Matches 10; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3 RAVRPALSGFGRVSG 19
DB 320 RAARPVCPGLDGEAGSG 336

RESULT 2

Q9U0Z1 PRELIMINARY; PRT; 2879 AA.
AC Q9U0Z1
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN L5893.06.
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Friedlin;
RA Murphy L., Harris D., Ivens A.C., Lawson D., Quail M.,
RA Rajandream M.A., Barrell B.G.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Friedlin;
RX MEDLINE=98146435; PubMed=9477341;
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome.";

RL Genome Res. 8:135-145(1998).
DR EMBL; AL117384; CAB5617.1; --
KW Hypothetical protein.
SQ SEQUENCE 2879 AA; 305038 MW; 1F75F7831C99B7DB CRC64;

Query Match 48.1%; Score 50; DB 5; Length 2879;
Best Local Similarity 52.6%; Pred. No. 1.2e+02;
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 QRAVRPALSGFGRVSGC 20
DB 1198 QRVLLPFGSHERRVGGC 1216

RESULT 3

Q8MT57 PRELIMINARY; PRT; 707 AA.
AC Q8MT57
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE REL1711p.
GN TRAP95 OR CG5465.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RA Stapleton M., Brokstein P., Hong L., Agbavani A., Carlson J.,
RA Champs M., Chavez C., Dorsett V., Dresnek D., Farfan D., Friese E.,
RA George K., Gonzalez M., Guarini H., Kronmiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleab J., Paragas V., Park S.,
RA Patel S., Phuananavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY118369; XAM48398.1; --
DR FlyBase; FBgn0034707; Trap95.
SQ SEQUENCE 707 AA; 77566 MW; 5620FD241C18B1FC CRC64;

Query Match 47.1%; Score 49; DB 5; Length 707;
Best Local Similarity 60.0%; Pred. No. 39;
Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 6 RPALSGFGRVSGC 20
DB 173 RPTLSGFGVASEGC 187

RESULT 4

Q9GYW8 PRELIMINARY; PRT; 818 AA.
AC Q9GYW8
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Thyroid hormone receptor-associated protein TRAP95.
GN TRAP95 OR CG5465.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Southworth J.W., Kennison J.A.;
RT "Transcriptional coactivators in Drosophila.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF289994; AAG02459.1; --
DR FlyBase; FBgn0034707; Trap95.
DR GO; GO:0004872; F:receptor activity; IEA.


```

KW Receptor.
SQ SEQUENCE 818 AA; 89974 MW; 6AC28B5D117C5678 CRC64;

Query Match
Best Local Similarity 47.1%; Score 49; DB 5; Length 818;
Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 6 RPALSGFGRVSGC 20
Db 173 RPTLSGFGVASEGC 187

RESULT 5
Q9W278 PRELIMINARY; PRT; 818 AA.
AC Q9W278;
DT 01-WAY-2000 (TrEMBLrel. 13, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-WAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CG5465 protein.
GN TRAP95 OR CG5465.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
RA April J.F., Agbayani A., An H.-J., Andrews-Frannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Bottier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA De Fabios B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kianos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busan D.A.,

```

```

RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.B., Doyle C., Dreenek D., Farfan D.,
RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleab J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Korman B., Carlson J.W., Celniker S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of Drosophila melanogaster genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF003456; AAF46815.2; -.
DR FlyBase; FBgn0034707; Trap95.
SQ SEQUENCE 818 AA; 89979 MW; 9AC366D99948F4D5 CRC64;

```

```

Query Match 47.1%; Score 49; DB 5; Length 818;
Best Local Similarity 60.0%; Pred. No. 46;
Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

```

QY 6 RPALSGFGRVSGC 20
Db 173 RPTLSGFGVASEGC 187

```

RESULT 6

```

Q91EQ9 PRELIMINARY; PRT; 150 AA.
AC Q91EQ9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Coat protein (Fragment).
GN CP.
OS Diocorea dumentorum virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potyviridae;
OC Potyvirus.
OX NCBI_TaxID=145665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ddBan60;
RA Mumford R.A., Jayaratne D.L., Boonham N., Canning E.C., Seal S.E.;
RT "The use of degenerate RT-PCR and sequencing to investigate the
RT diversity of potyviruses in yam (Dioscorea spp.)."
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE POTYVIRUSES POLYPROTEIN FAMILY.
DR EMBL; AJ054432; CAC43176.1; -.
DR GO; GO:0019028; C:Viral capsid; IEA.
DR GO; GO:0005198; F:Structural molecule activity; IEA.
DR InterPro; IPR001592; Poty coat.
DR Pfam; PF00767; Poty_coat_1.
DR Coat protein. 1
FT NON_TER 1
SQ SEQUENCE 150 AA; 17102 MW; A1B23B2638403518 CRC64;

```

```

Query Match          46.2%; Score 48; DB 12; Length 150;
Best Local Similarity 56.2%; Pred. No. 11;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 RAVRPALSGFDRGVGS 18
   | : | | | | | | | |
Db 109 RGAQPRFLGLDGKVG 124

RESULT 7
Q91EQ8 PRELIMINARY; PRT; 150 AA.
AC Q91EQ8;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Coat protein (Fragment).
GN CP.
OS Diocorea dumentorum virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potyviridae;
OC Potyvirus.
OX NCBI_TaxID=145665;
RN [1] SEQUENCE FROM N.A.
RP STRAIN=ddSLK19;
RA Mumford R.A., Jayaratne D.L., Boonham N., Canning E.C., Seal S.E.;
RT "The use of degenerate RT-PCR and sequencing to investigate the
RT diversity of potyviruses in yam (Dioscorea spp.).";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE POTYVIRUSES POLYPROTEIN FAMILY.
DR EMBL; AJ305433; CAC4317.1; -
DR GO; GO:0019028; C:Viral capsid; IEA.
DR GO; GO:0005198; F:Structural molecule activity; IEA.
DR InterPro; IPR001592; Poty_coat.
DR Pfam; PF00767; Poty_coat; 1.
KW Coat protein.
FT NON_TER
FT SEQUENCE 150 AA; 17103 MW; A1B23B2630E23518 CRC64;

Query Match          46.2%; Score 48; DB 12; Length 150;
Best Local Similarity 56.2%; Pred. No. 11;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 RAVRPALSGFDRGVGS 18
   | : | | | | | | | |
Db 109 RGAQPRFLGLDGKVG 124

RESULT 8
Q9RYP2 PRELIMINARY; PRT; 376 AA.
AC Q9RYP2;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Adenine deaminase-related protein.
GN DRA0268.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1] SEQUENCE FROM N.A.
RP STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Bisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodean R.J., Haft D.H., Gwinn W.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;

```

```

RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577 (1999).
DR EMBL; AE001863; AAF12376.1; -.
DR C75580; C75580.
DR TIGR; DRA0268; -.
KW Complete proteome.
SQ SEQUENCE 376 AA; 39845 MW; 7AB7FF32F8C45651 CRC64;

Query Match          46.2%; Score 48; DB 16; Length 376;
Best Local Similarity 64.7%; Pred. No. 29;
Matches 11; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 RAVRPALSGFDRGVGS 19
   | | | | | | | | | |
Db 275 RAVAPALRGSDRFPASG 291

RESULT 9
Q845V3 PRELIMINARY; PRT; 438 AA.
AC Q845V3;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Histidinol dehydrogenase.
GN HISD.
OS Burkholderia multivorans.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxID=87883;
RN [1] SEQUENCE FROM N.A.
RP STRAIN=ATCC 17616;
RA Komatsu H., Imura Y., Ohori A., Nagata Y., Tsuda M.;
RT "Distribution and Organization of Auxotrophic Genes on Multi-
RT Chromosomal Genome of Burkholderia multivorans ATCC17616.";
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB091436; BAC65270.1; -
DR GO; GO:0004399; F:histidinol dehydrogenase activity; IEA.
DR GO; GO:0000105; F:histidine biosynthesis; IEA.
DR InterPro; IPR001692; Histidinol_dh.
DR Pfam; PF00815; Histidinol_dh; 1.
DR PRINTS; PR00083; HOLDHDRGNASE.
DR ProDom; PD002680; Histidinol_dh; 1.
DR TIGRFAMs; TIGR00069; hisD; 1.
SQ SEQUENCE 438 AA; 46606 MW; 6D383BE84CDB5E25 CRC64;

Query Match          46.2%; Score 48; DB 2; Length 438;
Best Local Similarity 47.8%; Pred. No. 34;
Matches 11; Conservative 3; Mismatches 5; Indels 4; Gaps 1;

QY 2 QAVRPALSGFD----GRVSGC 20
   | | | | | | | | | |
Db 74 QDALQALDGLPEKARGRAGSGC 96

RESULT 10
Q98NT3 PRELIMINARY; PRT; 191 AA.
AC Q98NT3;
DT 01-OCT-2001 (TReMBLrel. 18, Created)
DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Hypothetical protein mlr9752.
GN MLR9752.
OS Rhizobium loti (Mesorhizobium loti).
OG Plasmid pMlb.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1] SEQUENCE FROM N.A.
RP

```

```

RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL; AP003017; BAB54922.1; -.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR InterPro; IPR008565; DUF847.
DR Pfam; PF05838; DUF847; 1.
KW Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 191 AA; 20104 MW; 8CD9D278BE1345AB CRC64;

Query Match 45.7%; Score 47.5; DB 16; Length 191;
Best Local Similarity 55.0%; Pred. No. 17;
Matches 11; Conservative 4; Mismatches 4; Indels 1; Gaps 1;

QY 1 LQAVRPALSG-FDGRVSG 19
Db ||:||||:|:|:|
12 LQALRPAYTGRIDGVLGMG 31

RESULT 11
Q984E8 PRELIMINARY; PRT; 283 AA.
AC Q984E8
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DE Hypothetical protein mlr8035.
GN Mlr8035.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL; AP003013; BAB53682.1; -.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR008565; DUF847.
DR InterPro; IPR000005; HTHArac.
DR Pfam; PF05838; DUF847; 1.
DR PROSITE; PS00041; HTH_ARAC_FAMILY_1; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 283 AA; 29859 MW; 697FE2B8A6F69ED1 CRC64;

Query Match 45.7%; Score 47.5; DB 16; Length 283;
Best Local Similarity 55.0%; Pred. No. 25;
Matches 11; Conservative 4; Mismatches 4; Indels 1; Gaps 1;

QY 1 LQAVRPALSG-FDGRVSG 19
Db ||:||||:|:|:|
103 LQALRPAYTGRIDGVLGMG 122

RESULT 12
Q982P9 PRELIMINARY; PRT; 283 AA.
AC Q982P9
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein mlr8547.
GN Mlr8547.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL; AP003014; BAB54407.1; -.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR008565; DUF847.
DR InterPro; IPR000005; HTHArac.
DR Pfam; PF05838; DUF847; 1.
DR PROSITE; PS00041; HTH_ARAC_FAMILY_1; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 283 AA; 29773 MW; ACS964A2A21B8A56 CRC64;

Query Match 45.7%; Score 47.5; DB 16; Length 283;
Best Local Similarity 55.0%; Pred. No. 25;
Matches 11; Conservative 4; Mismatches 4; Indels 1; Gaps 1;

QY 1 LQAVRPALSG-FDGRVSG 19
Db ||:||||:|:|:|
103 LQALRPAYTGRIDGVLGMG 122

RESULT 13
Q83A70 PRELIMINARY; PRT; 240 AA.
AC Q83A70
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE ABC transporter, ATP-binding protein.
GN CBU2038.
OS Coxiella burnetii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC Coxiellaceae; Coxiella.
OX NCBI_TaxID=777;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nine Mile phase I / RSA 493;
RX MEDLINE=22608657; PubMed=12704232;
RA Seshadri R., Paulsen I.T., Eisen J.A., Read T.D., Nelson K.E.,
RA Seshadri R., Paulsen I.T., Eisen J.A., Read T.D., Nelson K.E.,
RA Nelson W.C., Ward N.L., Tettelin H., Daviden T.M., Beanan M.J.,
RA DeBoy R.T., Daugherty S.C., Brinkac L.M., Madupu R., Dodson R.J.,
RA Khouri H.M., Lee K.H., Carty H.A., Scanlan D., Heinzen R.A.,
RA Thompson H.A., Samuel J.B., Fraser C.M., Heidelberg J.F.;
RT "Complete genome sequence of the Q-fever pathogen, Coxiella
RT burnetii.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:5455-5460(2003).
DR EMBL; AS016966; AAO91525.1; -.
DR TIGR; CBU2038; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.

```

```

DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Complete proteome.
SQ SEQUENCE 240 AA; 26743 MW; 3BBF12F4751AC133 CRC64;

Query Match 45.2%; Score 47; DB 16; Length 240;
Best Local Similarity 45.0%; Pred. No. 26;
Matches 9; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 1 LQRAVRPALSGFDGRVSGC 20
DB 131 LSELIWPLNGFDFTVGGGC 150

RESULT 14
Q8WSC0 PRELIMINARY; PRT; 166 AA.
AC Q8WSC0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Putative GTPase regulator protein.
GN OSUNB0013K08.11.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Buell C.R., Yuan O., Qiyang S., Liu J., Moffat K.S., Hill J.N.,
RA Ganseberger K., Brenner M., Burgess S., Hance M., Shvartsbeyn M.,
RA Taitrin T., Riggs F., Hsiao J., Ziemann V., Blunt S., Pai G.,
RA Vanaken S.E., Utterback T.R., Feldblyum T.V., Kalb E., Quackenbush J.,
RA Salzberg S.L., White O., Fraser C.M.;
RT "Oryza sativa chromosome 3 BAC OSUNB0013K08 genomic sequence.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC092390; AL311680.1; -.
DR Gramene; Q8WSC0; -.
SQ SEQUENCE 166 AA; 18447 MW; 76A28ED4DEEE33C3 CRC64;

Query Match 44.2%; Score 46; DB 10; Length 166;
Best Local Similarity 50.0%; Pred. No. 25;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 QRAVRPALSGFDGRVSGG 19
DB 89 EKAARAARKGFDGGSGEG 106

RESULT 15
Q8EHA4 PRELIMINARY; PRT; 1482 AA.
AC Q8EHA4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Glutamate synthase, large subunit.
GN GLTB OR SOI325.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NR-1;
RX MEDLINE=22297686; PubMed=12368813;

RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A., C.M.;
RA Feldblyum T.V., Smith H.O., Venter J.C., Nealon K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis.",
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AE015576; AAN54390.1; -.
DR TIGR; SO1325; -.
DR GO; GO:0015930; F:glutamate synthase activity; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006537; P:glutamate biosynthesis; IEA.
DR GO; GO:0006807; P:nitrogen metabolism; IEA.
DR InterPro; IPR002198; ADH_short.
DR InterPro; IPR002489; DUF14.
DR InterPro; IPR003009; FMN_enzyme.
DR InterPro; IPR002932; Glu_synthase.
DR InterPro; IPR006982; Glu_synth_centr.
DR InterPro; IPR006981; Glu_synth_NTN.
DR Pfam; PF01645; Glu_synthase; 1.
DR Pfam; PF04897; Glu_synth_NTN; 1.
DR Pfam; PF04898; Glu_syn_central; 1.
DR Pfam; PF01493; GXGXG; 1.
DR PROSITE; PS00061; ADH_SHORT; 1.
KW Complete proteome.
SQ SEQUENCE 1482 AA; 163011 MW; 10FDB14E58BF474 CRC64;

Query Match 43.8%; Score 45.5; DB 16; Length 1482;
Best Local Similarity 37.9%; Pred. No. 3.2e+02;
Matches 11; Conservative 1; Mismatches 6; Indels 11; Gaps 1;

QY 3 RAVRPALSGF-----DGRVSGC 20
DB 28 RIVRTAIHGLDRMKHKGIIASDGRGTGDC 56

RESULT 16
Q83611 PRELIMINARY; PRT; 287 AA.
AC Q83611;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Phosphatidate cytidyltransferase (CDSA).
GN TP0602.
OS Treponema pallidum.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nichols;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete.";
RL Science 281:375-388(1998).
DR EMBL; AE001235; AAC65575.1; -.
DR PIR; D71304; D71304.
DR TIGR; TP0602; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004605; F:phosphatidate cytidyltransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008654; P:phospholipid biosynthesis; IEA.

```

```
DR InterPro; IPR000374; PC trans.
DR Pfam; PF01148; CTP transf 1; 1.
KW Transferrase; Nucleotidyltransferase; Complete proteome.
SQ SEQUENCE 287 AA; 30866 MW; 42608D3A0C1B0FB CRC64;

Query Match 43.3%; Score 45; DB 16; Length 287;
Best Local Similarity 36.8%; Pred. No. 65;
Matches 7; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 2 QRAVRPALSGFGRVSGC 20
    : : : : :
Db 186 KKSIAGFIGGFAAGVAGC 204
    : : : : :

RESULT 17
Q98KT8 PRELIMINARY; PRT; 299 AA.
AC Q98KT8;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Succinyl-CoA synthetase alpha subunit.
GN MLR1326.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Iidesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL; AF002997; BAB48726.1; -.
DR GO; GO:0003824; P:catalytic activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR003781; CoA binding.
DR InterPro; IPR005811; CoA ligase.
DR InterPro; IPR005810; CoA lig alpha.
DR Pfam; PF02629; CoA binding; 1.
DR Pfam; PF00549; ligase-CoA; 1.
DR PRINTS; PR01798; SCOASYNTHASE.
DR TIGRFAMs; TIGR01019; succoAalpha; 1.
DR PROSITE; PS00399; SUCCINYL_COA_LIG_2; 1.
KW Complete proteome.
SQ SEQUENCE 299 AA; 31601 MW; 78194A84A92079FF CRC64;

Query Match 43.3%; Score 45; DB 16; Length 299;
Best Local Similarity 44.4%; Pred. No. 68;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 LQRAVRPALSGFGRVGS 18
    : : : : :
Db 5 LNRSTRIVQGFYKIGS 22
    : : : : :

RESULT 18
Q7XXP6 PRELIMINARY; PRT; 373 AA.
AC Q7XXP6;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Callus;
RA Moriguchi K., Ito Y., Yamazaki Y., Kurata N.;
RT "Finding of various plant nuclear proteins using yeast nuclear
RT transportation trap system - a proteomal approach.";
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB110206; BAC78598.1; -.
KW Hypothetical protein.
SQ SEQUENCE 373 AA; 37513 MW; 9B622F770CFC6DA7 CRC64;

Query Match 43.3%; Score 45; DB 10; Length 373;
Best Local Similarity 61.5%; Pred. No. 86;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 8 ALSGFDGRVSGC 20
    : : : : :
Db 255 ALAGSDGRIVGCG 267
    : : : : :

RESULT 19
Q7XM00 PRELIMINARY; PRT; 379 AA.
AC Q7XM00;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE OSJNB0086006.9 protein.
GN OSJNB0086006.9.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA Han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X.,
RA Liu Y.L., Mu J., Yu Z., Chen L., Fan D.L., Weng Q.J., Zhang L.,
RA Lu Y.Q., Yu S.L., Liu X.H., Lu T.T., Zhang Y.J., Lu Y., Li C., Li T.,
RA Zhang Y., Hu H., Jia P.X., Qian Y.M., Ying K., Zhou B., Chen Z.H.,
RA Hao P., Zhang L., Wu M., Zhang R.Q., Guan J.P., Fu G., Wang S.Y.,
RA Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F., Jia J., Yin H.F.,
RA Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y., Shao C.Y., Sun Y.,
RA Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W., Sheng H.H.,
RA Gu J.L., Chen S.T., Ni L., Zhu P.H., Hong G.F.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL662981; CAE04861.1; -.
SQ SEQUENCE 379 AA; 38241 MW; FC493F8D769E4670 CRC64;

Query Match 43.3%; Score 45; DB 10; Length 379;
Best Local Similarity 61.5%; Pred. No. 88;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 8 ALSGFDGRVSGC 20
    : : : : :
Db 255 ALAGSDGRIVGCG 267
    : : : : :

RESULT 20
Q9U937 PRELIMINARY; PRT; 398 AA.
AC Q9U937;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE RNA helicase-1.
OS Plasmodium cynomolgi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5827;
RN [1]
RP SEQUENCE FROM N.A.
```

RA Song P., Malhotra P.;
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 CC 1- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY.
 DR EMBL; AJ245475; CAB51741.1; -.
 DR HSSP; Q58083; IHV8.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO; GO:0016787; F:hydrolyase activity; IEA.
 DR GO; GO:0003676; F:nucleic acid binding; IEA.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR000629; DEAD_box.
 DR InterPro; IPR001650; Helicase_C.
 DR Pfam; PF00270; DEAD; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR SMART; SM00490; HELIC; 1.
 DR PROSITE; PS00039; DEAD_ATP_HELICASE; 1.
 KW ATP-binding; Helicase; Hydrolase.
 SQ SEQUENCE 398 AA; 45342 MW; 357FE8889129589 CRC64;

Query Match 43.3%; Score 45; DB 5; Length 398;
 Best Local Similarity 45.0%; Pred. No. 92;
 Matches 9; Conservative 5; Mismatches 4; Indels 2; Gaps 1;

QY 2 QRAVRPALSGFD--GRVSG 19
 DB 51 QRGKIPILNGYDTIGQAQSG 70

RESULT 21

Q8IKF0 PRELIMINARY; PRT; 398 AA.
 AC Q8IKF0;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-WAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE RNA helicase-1, putative.
 GN PF14_0655.
 OS Plasmodium falciparum (isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=36329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=3D7.
 RX MEDLINE=22255705; PubMed=12368864;
 RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
 RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
 RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
 RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
 RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
 RA Martin D.M.A., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
 RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
 RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
 RA Fraser C.M., Barrell B.;
 RT "Genome sequence of the human malaria parasite Plasmodium falciparum";
 RL Nature 419:498-511(2002).
 DR EMBL; AS014827; AAN37268.1; -.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.
 DR GO; GO:0003676; F:nucleic acid binding; IEA.
 DR InterPro; IPR001410; DEAD.
 DR InterPro; IPR000629; DEAD_box.
 DR InterPro; IPR001650; Helicase_C.
 DR Pfam; PF00270; DEAD; 1.
 DR Pfam; PF00271; helicase_C; 1.
 DR SMART; SM00487; DEXDC; 1.
 DR SMART; SM00490; HELIC; 1.
 DR PROSITE; PS00039; DEAD_ATP_HELICASE; 1.
 KW Helicase.
 SQ SEQUENCE 398 AA; 45310 MW; 701B73A48C5F3476 CRC64;
 Query Match 43.3%; Score 45; DB 5; Length 398;

Best Local Similarity 45.0%; Pred. No. 92;
 Matches 9; Conservative 5; Mismatches 4; Indels 2; Gaps 1;
 QY 2 QRAVRPALSGFD--GRVSG 19
 DB 51 QRGKIPILNGYDTIGQAQSG 70

RESULT 22

Q9AYM0 PRELIMINARY; PRT; 405 AA.
 AC Q9AYM0;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative AT-hook DNA-binding protein.
 GN OSJNBA0003019.1.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaeae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Buell C.R., Yuan Q., Moffat K.S., Hill J.N., Burr P.C., Hsiao J.,
 RA Zismann V., Pai G., Bowman C.L., Fujii C.Y., VanAken S.E.,
 RA Bowman C.L., Craven B., Utterback T.R., Khatai H., Feldblyum T.V.,
 RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
 RT "Oryza sativa chromosome 10 BAC OSJNBA0003019 genomic sequence."
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC060755; AAK00433.1; -.
 DR Gramene; Q9AYM0; -.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR000637; AT_hook.
 DR InterPro; IPR005175; DUF296.
 DR Pfam; PF02178; AT_hook; 2.
 DR Pfam; PF03479; DUF296; 1.
 DR SMART; SM00384; AT_hook; 2.
 KW DNA-binding.
 SQ SEQUENCE 405 AA; 41399 MW; 274994FD12D49816 CRC64;

Query Match 43.3%; Score 45; DB 10; Length 405;
 Best Local Similarity 69.2%; Pred. No. 94;
 Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 8 ALSGPDGRVSGC 20
 DB 273 ALAGSDGRVLGOC 285

RESULT 23

Q7X976 PRELIMINARY; PRT; 405 AA.
 AC Q7X976;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative AT-hook DNA-binding protein.
 GN OSJNBA0003019.1.
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaeae; Oryza.
 OX NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA The Rice Chromosome 10 Sequencing Consortium;
 RT "In-depth view of structure, activity, and evolution of rice chromosome 10.";

```
RL Science 300:1566-1569 (2003).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Buell C.R., Wing R.A., McCombie W.R., Messing J., Yuan Q.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AEO17121; AAP55117.1; -.
KW DNA-binding.
SQ SEQUENCE 405 AA; 41399 MW; 274994FD12D49816 CRC64;

Query Match 43.3%; Score 45; DB 10; Length 405;
Best Local Similarity 69.2%; Pred. No. 94;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 8 ALSGDFGRVSGC 20
Db ||:|||||
273 ALAGSDGRVLGGC 285

RESULT 24
Q9LKR6 PRELIMINARY; PRT; 453 AA.
AC Q9LKR6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE T26D3.3 protein.
GN T26D3.3.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
RN [1]
NCBI_TaxID=3702;
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA WashU;
RT "The A. thaliana Genome Sequencing Project.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Wilson R.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF262043; AAP88016.1; -.
DR HSSP; P18670; IJAC.
DR InterPro; IPR001229; Jacalin_lectin.
DR Pfam; PF01419; Jacalin; 3.
SQ SEQUENCE 453 AA; 50205 MW; D37A33D335371E172 CRC64;

Query Match 43.3%; Score 45; DB 10; Length 453;
Best Local Similarity 75.0%; Pred. No. 1.1e+02;
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 8 ALSGDFGRVSGC 19
Db ||:|||||
275 ALIGHFGRVAGC 286

RESULT 25
Q8LQF2 PRELIMINARY; PRT; 484 AA.
AC Q8LQF2;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE DNA-binding protein-like protein.
GN P0004D12.17.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
RN [1]
NCBI_TaxID=39947;

[1]
RN SEQUENCE FROM N.A.
RP STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
clone:P0004D12.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
TRANSCRIPTION FACTORS.
DR EMBL; AF003433; BAB92673.1; -.
DR Gramene; Q8LQF2; -.
DR GO; GO:0003677; F.DNA binding; IEA.
DR InterPro; IPR001092; HLH_basic.
DR Pfam; PF00010; HLH; 1.
DR SMART; SM00353; HLH; 1.
DR PROSITE; PS00888; HLH_2; 1.
KW DNA-binding.
SQ SEQUENCE 484 AA; 50351 MW; A0C5FE7D6B6E86FDCRC64;

Query Match 43.3%; Score 45; DB 10; Length 484;
Best Local Similarity 61.1%; Pred. No. 1.1e+02;
Matches 11; Conservative 1; Mismatches 4; Indels 2; Gaps 1;

QY 2 QRAVRPALSGFGRVSGC 19
Db ||:|||||
125 ERAAR--LSGFDARGGGC 140

RESULT 26
Q8XDM4 PRELIMINARY; PRT; 484 AA.
AC Q8XDM4;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Xylulokinase.
GN XYL OR Z4989 OR ECS4447.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoukis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22 (2001).
DR EMBL; AS005583; AAG58713.1; -.
DR EMBL; AF002565; BAB37870.1; -.
DR PIR; E86031; E86031.
DR PIR; G91184; G91184.
DR GO; GO:0016301; P.kinase activity; IEA.
DR GO; GO:0004856; F.xylulokinase activity; IEA.
DR GO; GO:0005975; P.carbohydrate metabolism; IEA.
DR GO; GO:0005997; P.xylulose metabolism; IEA.
DR InterPro; IPR000577; FGGY_kin.
```

```

DR InterPro; IPR006000; Xylulokinase.
DR Pfam; PF00370; FGGY; 1.
DR Pfam; PF02782; FGGY_C; 1.
DR TIGRFAMs; TIGR01312; XylB; 1.
DR PROSITE; PS00933; FGGY_KINASES_1; 1.
DR PROSITE; PS00445; FGGY_KINASES_2; 1.
KW Kinase; Complete proteome.
SQ SEQUENCE 484 AA; 52589 MW; 3E22A7C7465BAA89 CRC64;

Query Match 43.3%; Score 45; DB 16; Length 484;
Best Local Similarity 47.4%; Pred. No. 1.1e+02;
Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 2 QRAVRPALSGFDGRVSGC 20
DB 87 QRVLRPAILWNGRCQAQC 105

RESULT 27
Q8FCE4
ID Q8FCE4 PRELIMINARY; PRT; 484 AA.
AC Q8FCE4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Xylulose kinase (EC 2.7.1.17).
GN XylB OR C4384.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=217992;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=2238234; PubMed=12471157;
RX Wei J., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Welch R.A., Burland V., Plunkett G. III, S.-R., Boutin A., Hackett J., Stroud D.,
RA Rasko D., Buckles E.L., Liou S.-R., Zhou S., Schwartz D.C., Perna N.T.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Blattner F.R.;
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AS016768; AAN82820.1; -.
DR GO; GO:0016301; F-kinase activity; IEA.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0004856; F:xylulokinase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0005997; P:xylulose metabolism; IEA.
DR InterPro; IPR00577; FGGY kin.
DR InterPro; IPR006000; Xylulokinase.
DR Pfam; PF00370; FGGY; 1.
DR Pfam; PF02782; FGGY_C; 1.
DR TIGRFAMs; TIGR01312; XylB; 1.
DR PROSITE; PS00933; FGGY_KINASES_1; 1.
DR PROSITE; PS00445; FGGY_KINASES_2; 1.
KW Kinase; Transferase; Complete proteome.
SQ SEQUENCE 484 AA; 52627 MW; 716E8C6E14271E1 CRC64;

Query Match 43.3%; Score 45; DB 16; Length 484;
Best Local Similarity 47.4%; Pred. No. 1.1e+02;
Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 2 QRAVRPALSGFDGRVSGC 20
DB 87 QRVLRPAILWNGRCQAQC 105

RESULT 28
Q83J34
ID Q83J34 PRELIMINARY; PRT; 484 AA.
AC Q83J34;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)

```

```

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Xylulokinase.
GN XylB OR SE3608 OR S4161.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=301 / Serotype 2a;
RX MEDLINE=2272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Lu W., Wang J., Liu H.,
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
RA Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
through comparison with genomes of Escherichia coli K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=2457T / ATCC 700930 / Serotype 2a;
RX MEDLINE=22590274; PubMed=12704152;
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,
RA Schwartz D.C., Blattner F.R.;
RT "Complete genome sequence and comparative genomics of Shigella
flexneri serotype 2a strain 2457T.";
RL Infect. Immun. 71:2775-2786(2003).
DR EMBL; AE015369; AAN45058.1; -.
DR EMBL; AE016991; AAP19131.1; -.
DR GO; GO:0016301; F-kinase activity; IEA.
DR GO; GO:0004856; F:xylulokinase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0005997; P:xylulose metabolism; IEA.
DR InterPro; IPR00577; FGGY kin.
DR InterPro; IPR006000; Xylulokinase.
DR Pfam; PF00370; FGGY; 1.
DR Pfam; PF02782; FGGY_C; 1.
DR TIGRFAMs; TIGR01312; XylB; 1.
DR PROSITE; PS00933; FGGY_KINASES_1; 1.
DR PROSITE; PS00445; FGGY_KINASES_2; 1.
KW Kinase; Complete proteome.
SQ SEQUENCE 484 AA; 52650 MW; 5296AD46195A6DEA CRC64;

Query Match 43.3%; Score 45; DB 16; Length 484;
Best Local Similarity 47.4%; Pred. No. 1.1e+02;
Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 2 QRAVRPALSGFDGRVSGC 20
DB 87 QRVLRPAILWNGRCQAQC 105

RESULT 29
Q8EB00
ID Q8EB00 PRELIMINARY; PRT; 499 AA.
AC Q8EB00;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to Arabidopsis thaliana (Mouse-ear cress), nodulin /
DE Glutamate-ammonia ligase-like protein (EC 6.3.1.2) (Glutamine
DE synthetase) (Glutamate--ammonia ligase).
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RX MEDLINE=22092622; PubMed=12097910;
RA Gloeckner G., Eichinger L., Szafranski K., Fachebat J., Dear P.,

```



```

RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT "Sequence and analysis of chromosome 2 of Dictyostelium discoideum.";
RL Nature 418:79-85(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Baumgart C.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC117075; AAC50777.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004356; F:glutamate-ammonia ligase activity; IEA.
DR GO; GO:0016874; F:ligase activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0009399; F:nitrogen fixation; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008147; Gln synt beta.
DR InterPro; IPR008146; Gln synt_C.
DR InterPro; IPR00508; Peptidase_S26.
DR Pfam; PF00120; gln-synt; 1.
DR ProDom; PD001057; Gln synt_C; 1.
DR PROSITE; PS00761; SPASE_I_3; 1.
KW Ligase.
SQ SEQUENCE 499 AA; 56341 MW; 42A903AEE0AA6365 CRC64;

Query Match 43.3%; Score 45; DB 5; Length 499;
Best Local Similarity 88.9%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 12 FDGRVGGC 20
DB 293 FDGLVGGC 301

RESULT 30
Q7UWH2 PRELIMINARY; PRT; 1553 AA.
AC Q7UWH2;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN RB2038.
OS Rhodopirellula baltica.
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC Planctomycetaceae; Pirellula.
OX NCBI_TaxID=117;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1;
RX MEDLINE=22735913; PubMed=12835416;
RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,
RA Schleener H., Amann R., Reinhardt R.;
RT "Complete genome sequence of the marine planctomycete Pirellula sp.
RT strain 1.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).
DR EMBL; BX294136; CAD72391.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1553 AA; 172937 MW; D737F17A94BE7C0D CRC64;

Query Match 43.3%; Score 45; DB 16; Length 1553;
Best Local Similarity 64.3%; Pred. No. 4e+02;
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 7 PALSFGDGRVGGC 20
DB 1114 PQLRGFDARVLVGC 1127

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2004, 06:49:47 ; Search time 48 Seconds
(without alignments)
117.728 Million cell updates/sec

Title: 09549186-8

Perfect score: 101

Sequence: 1 LQRAVLALSGDFRVSQC 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : A Geneseq_29Jan04:*

1: Geneseqp1980s:*

2: Geneseqp1990s:*

3: Geneseqp2000s:*

4: Geneseqp2001s:*

5: Geneseqp2002s:*

6: Geneseqp2003as:*

7: Geneseqp2003bs:*

8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	89	88.1	20	ADB87347	Adb87347 Vaccinatio
2	76	75.2	19	AAR45910	Aar45910 Foot and
3	47	46.5	359	AAU33620	Aau33620 Pseudomon
4	47	46.5	359	ABU15565	Abu15565 Protein e
5	46	45.5	339	ADC94164	Adc94164 E. faeciu
6	45.5	45.0	229	AAG92858	Aag92858 C. glutami
7	45	44.6	263	ABB48000	Abb48000 Listeria
8	45	44.6	263	ABU32866	Abu32866 Protein e
9	44	43.6	233	ABM67849	Abm67849 Phototrab
10	43.5	43.1	53	AAU44797	Aau44797 Propionib
11	43.5	43.1	53	ABM41316	Abm41316 Propionib
12	43	42.6	141	AAU63931	Aau63931 Propionib
13	43	42.6	141	ABM60450	Abm60450 Propionib
14	43	42.6	228	AAR98742	Aar98742 Tazaroeten
15	43	42.6	287	ABU48621	Abu48621 Protein e
16	43	42.6	352	ABBA48949	Abba48949 Listeria
17	43	42.6	586	ABBA7493	Abba7493 Drosophil
18	43	42.6	1189	ABB68757	Abb68757 Drosophil
19	43	42.6	1221	ABU38466	Abu38466 Protein e
20	42.5	42.1	78	AAU92483	Aau92483 Human dig
21	42.5	42.1	78	AAU22528	Aau22528 Novel hum
22	42.5	42.1	78	ADB32368	Adb32368 Human nov
23	42	41.6	17	AAU00762	Aau00762 Human bon
24	42	41.6	133	AAU44009	Aau44009 Propionib
25	42	41.6	133	ABM40528	Abm40528 Propionib

26	42	41.6	133	7	ADC07742	Adc07742 Rice prot
27	42	41.6	155	6	ADA36132	Ada36132 Acinetoba
28	42	41.6	346	4	ABG06006	Abg06006 Novel hum
29	42	41.6	394	4	ABB59456	Abb59456 Drosophil
30	42	41.6	466	5	ABB92610	Abb92610 Herbicida
31	42	41.6	799	5	ABF65781	Abf65781 Bifidobac
32	42	41.6	898	5	ABG70103	Abg70103 Human pre
33	42	41.6	943	5	ABG91056	Abg91056 Neisseria
34	42	41.6	1037	4	ABB65026	Abb65026 Drosophil
35	42	41.6	1315	6	ABM65844	Abm65844 Propionib
36	42	41.6	1872	2	AAR56493	Aar56493 TATA-bind
37	42	41.6	1872	2	AAW06078	Aaw06078 Drosophil
38	42	41.6	1872	2	AAW25030	Aaw25030 TATA-bind
39	42	41.6	1893	2	AAR56491	Aar56491 TATA-bind
40	42	41.6	1893	2	AAW06082	Aaw06082 Human TAT
41	42	41.6	1893	2	AAW25020	Aaw25020 TATA-bind
42	42	41.6	1924	4	ABG06008	Abg06008 Novel hum
43	41.5	41.1	471	3	ABM15927	Abm15927 E. coli p
44	41.5	41.1	471	4	AAG98974	Aag98974 E. coli g
45	41.5	41.1	620	3	AAG31409	Aag31409 Arabidops
46	41.5	41.1	649	3	AAG31408	Aag31408 Arabidops
47	41.5	41.1	678	3	AAG31682	Aag31682 Arabidops
48	41.5	41.1	689	3	AAG31407	Aag31407 Arabidops
49	41.5	41.1	689	5	ABB93078	Abb93078 Herbicida
50	41.5	41.1	707	3	AAG31681	Aag31681 Arabidops
51	41.5	41.1	747	3	AAG31680	Aag31680 Arabidops
52	41	40.6	37	3	ABM20688	Abm20688 Polymeric
53	41	40.6	37	6	ABP80484	Abp80484 N. gonorr
54	41	40.6	37	6	ABP77386	Abp77386 N. gonorr
55	41	40.6	56	4	AAU41641	Aau41641 Propionib
56	41	40.6	56	6	ABM38160	Abm38160 Propionib
57	41	40.6	67	4	AAU43462	Aau43462 Propionib
58	41	40.6	67	6	ABM39981	Abm39981 Propionib
59	41	40.6	73	2	AAV22514	Aay22514 Xenopus h
60	41	40.6	99	7	ADB74327	Adb74327 Mycobacte
61	41	40.6	115	7	ADB74546	Adb74546 Mycobacte
62	41	40.6	127	6	ABP80865	Abp80865 N. gonorr
63	41	40.6	127	6	ABP80891	Abp80891 N. gonorr
64	41	40.6	127	6	ABP80485	Abp80485 N. gonorr
65	41	40.6	127	6	ABP77395	Abp77395 N. gonorr
66	41	40.6	127	6	ABP79509	Abp79509 N. gonorr
67	41	40.6	161	6	ABP77628	Abp77628 N. gonorr
68	41	40.6	208	7	ADC87091	Adc87091 Human GPC
69	41	40.6	215	3	AAG46917	Aag46917 Arabidops
70	41	40.6	255	6	ABU44396	Abu44396 Protein e
71	41	40.6	258	3	AAG46916	Aag46916 Arabidops
72	41	40.6	259	6	ABU14599	Abu14599 Protein e
73	41	40.6	263	4	AAU33415	Aau33415 Enterococ
74	41	40.6	263	4	AAU35094	Aau35094 Enterococ
75	41	40.6	297	3	AAG46915	Aag46915 Arabidops
76	41	40.6	378	4	ABG93428	Abg93428 Human pro
77	41	40.6	378	4	AAG89286	Aag89286 Human sec
78	41	40.6	378	5	ABB97425	Abb97425 Novel hum
79	41	40.6	454	4	AAG91457	Aag91457 C. glutami
80	41	40.6	502	6	ABU43877	Abu43877 Protein e
81	40.5	40.1	334	6	AAU36314	Aau36314 Pseudomon
82	40.5	40.1	334	6	ABU38537	Abu38537 Protein e
83	40	39.6	85	3	ABM56505	Abm56505 Human pro
84	40	39.6	102	3	AAG25014	Aag25014 Arabidops
85	40	39.6	202	4	AAG92383	Aag92383 C. glutami
86	40	39.6	203	4	ABG22564	Abg22564 Novel hum
87	40	39.6	220	4	ABG29258	Abg29258 Novel hum
88	40	39.6	229	4	AAG98977	Aag98977 E. coli g
89	40	39.6	340	3	AAG51488	Aag51488 Arabidops
90	40	39.6	347	4	ABG23488	Abg23488 Novel hum
91	40	39.6	349	4	ABG11319	Abg11319 Novel hum
92	40	39.6	357	6	ABU22779	Abu22779 Protein e
93	40	39.6	386	4	AAE02551	Aae02551 A. thalia
94	40	39.6	386	7	ADD30730	Add30730 Plant yie
95	40	39.6	433	4	AAV72079	Aay72079 Nicotiana
96	40	39.6	433	4	AAV72080	Aay72080 Nicotiana
97	40	39.6	439	4	ABG17899	Abg17899 Novel hum
98	40	39.6	441	3	AAG51487	Aag51487 Arabidops

99 40 39.6 443 4 ABB64394 Drosophil
100 40 39.6 444 3 AAG51486 Arabidops

Abb64394 Drosophil
Aag51486 Arabidops

ALIGNMENTS

RESULT 1

ADB87347
ID ADB87347 standard; peptide; 20 AA.

XX AC ADB87347;
XX 04-DEC-2003 (first entry)
DT XX Vaccination related retro-partly inverso peptide #1.
DE XX immunoretroid; anti-immunoretroid; CONH linkage; NHCO linkage;
KW retropeptide; retroinverso peptide; vaccine; viral; bacterial infection;
KW autoimmune disease; neurodegenerative disease; retro-partly;
KW inverso peptide.
XX OS Unidentified.

XX Key Location/Qualifiers
FT Modified-site 1 /note= "Modified by OH-m. Peptide can be an R or S
FT isomer"

FT Misc-difference 2 /note= "D-form residue"
FT Misc-difference 3 /note= "D-form residue"
FT Misc-difference 4 /note= "D-form residue"
FT Misc-difference 5 /note= "D-form residue"
FT Misc-difference 6 /note= "D-form residue"
FT Misc-difference 7 /note= "D-form residue"
FT Misc-difference 8 /note= "D-form residue"
FT Misc-difference 9 /note= "D-form residue"
FT Misc-difference 10 /note= "D-form residue"
FT Misc-difference 11 /note= "D-form residue"
FT Misc-difference 12 /note= "D-form residue"
FT Misc-difference 13 /note= "D-form residue"
FT Misc-difference 14 /note= "D-form residue"
FT Misc-difference 15 /note= "D-form residue"
FT Misc-difference 16 /note= "D-form residue"
FT Misc-difference 17 /note= "D-form residue"
FT Misc-difference 18 /note= "D-form residue"
FT Modified-site 18 /note= "C-terminal amide"

PN FR2717081-A1.

XX 15-SEP-1995.

XX 14-MAR-1994; 94FR-00002950.

XX 14-MAR-1994; 94FR-00002950.

XX (CNRS) CENT NAT RECH SCI.

XX Guichard G, Muller S, Briand J, Regenmortel MHV;

DR WPI; 1995-322414/42.

XX Therapeutic and diagnostic uses of retro peptide analogues - corresp. to
PT parent peptide chains with CONH linkages replaced by NHCO linkages, also
PT antibodies against the peptide(s).

PS Disclosure; Page 21; 58pp; French.

XX This invention relates to the novel uses of 'immunoretroids' or anti-
CC immunoretroid antibodies, where the immunoretroids are peptide analogues
CC in which one or more (preferably all) of the CONH linkages in the chain
CC of the corresponding parent peptides are replaced by NHCO linkages and
CC the chirality of each amino acid residue, whether involved in NHCO
CC linkages or not, is either conserved or inverted with regards to the
CC corresponding amino acid residue in the parent peptides. For example,
CC 'retropeptides' or 'retroinverso peptides', provided that the
CC immunoretroids are capable of forming complexes with the anti-
CC immunoretroid antibodies and with antibodies directed against the parent
CC peptides or parent proteins and/or the parent peptide enantiomers or
CC parent protein enantiomers. The immunoretroids are used to prepare
CC medicaments for preventing or treating pathologies associated with the
CC presence of an exogenous or endogenous protein capable of being
CC implicated directly or indirectly in the appearance and/or development of
CC the pathologies. Immunoretroids can also be used to prepare vaccines for
CC preventing pathologies associated with the presence of an exogenous or
CC endogenous protein recognised by antibodies directed against
CC immunoretroids. Comparisons containing immunoretroids associated with a
CC carrier molecule capable of inducing production of antibodies against an
CC exogenous or endogenous protein responsible for a pathology, or of
CC inducing a cytotoxic cellular immune response are useful as vaccines.
CC Pathologies that can be diagnosed or treated are especially viral or
CC bacterial infections, autoimmune diseases and neurodegenerative diseases.
CC This sequence represents a vaccination related retro-partly inverso
XX peptide relating to the retropeptides of the invention.

XX Sequence 20 AA;

Query Match 88.1%; Score 89; DB 2; Length 20;
Best Local Similarity 90.0%; Pred. No. 2.3e-07;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 LQRAVRLALSGFDGRVSGSC 20
| | | | | | | | | | | | | | | | | |
DB 1 LQRAVRLALSGFDGRVSGSC 20

RESULT 2

AAR45910
ID AAR45910 standard; peptide; 19 AA.

XX AAR45910;

XX 27-AUG-2003 (revised)

DT 25-MAR-2003 (revised)

DT 27-JUL-1994 (first entry)

DE Foot and Mouth Disease Virus immunogenic peptide P6.

XX immunodominant surface-exposed epitope; T-cell proliferation;
KW stimulation; immunogenicity; increase; enhance; vaccine;
KW Foot and Mouth Disease Virus; FMDV; Aphthovirus; LCP; lipid anchor;
KW immunogenic carrier.

XX Foot-and-mouth disease virus.

XX WO9402506-A1.

XX 03-FEB-1994.

XX 23-JUL-1993; 93WO-CB001558.

XX 24-JUL-1992; 92GB-00015780.

PA (UNLO) UNIV LONDON SCHOOL PHARMACY.
 XX Toth I, Gibbons WA;
 XX WPI; 1994-048791/06.
 XX New lipidic amino acid based anchor system - for attachment of short
 PT synthetic peptide(s) to enhance their antigenicity and for use, e.g., as
 PT vaccines.
 XX
 PS Example 3; Page 36; 50pp; English.
 XX
 CC A lipidic amino acid based anchor system was synthesised. Eight copies of
 CC the FMDV immunogenic peptide P6 (AAR45910) were attached to the anchor to
 CC give (peptide)8lys4lys2lys(HNCH(CH2)13Me)(CO)3NH2. When injected into
 CC cows, immunogenicity was found to be 10 times higher than would be
 CC expected from a conventional BSA- or KLH-conjugated FMDV. (Updated on 25-
 CC MAR-2003 to correct PN field.) (Updated on 27-AUG-2003 to correct OS
 CC field.)
 XX
 SQ Sequence 19 AA;
 Query Match 75.2%; Score 76; DB 2; Length 19;
 Best Local Similarity 94.1%; Pred. No. 2.8e-05;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 3 RAVRLALSGFDGRVGS 19
 |||||
 DB 3 RAVRPALSGFDGRVGS 19
 |||||
 RESULT 3
 AAU33620
 ID AAU33620 standard; protein; 359 AA.
 XX
 AC AAU33620;
 XX
 DT 14-FEB-2002 (first entry)
 XX
 DE Pseudomonas aeruginosa cellular proliferation protein #64.
 XX
 KW Antisense; prokaryotic cellular proliferation protein; antibiotic;
 KW antibacterial; drug design.
 XX
 OS Pseudomonas aeruginosa.
 XX
 PN WO200170955-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 21-MAR-2001; 2001WO-US009180.
 XX
 PR 21-MAR-2000; 2000US-0191078P.
 PR 23-MAY-2000; 2000US-0206848P.
 PR 26-MAY-2000; 2000US-0207727P.
 PR 23-OCT-2000; 2000US-0242578P.
 PR 27-NOV-2000; 2000US-0253625P.
 PR 22-DEC-2000; 2000US-0257931P.
 PR 16-FEB-2001; 2001US-0269308P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 XX
 DR WPI; 2001-611495/70.
 DR N-PSDB; AAS51479.
 XX
 DR New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids.
 XX
 PS Example 3; SEQ ID NO 5116; 511pp; English.
 XX

CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the genes,
 CC their use in the discovery of novel antibiotics, the essential genes
 CC themselves and the encoded proteins. The prokaryotes used are Escherichia
 CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
 CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
 CC useful for the identification of potential new targets for antibiotic
 CC development. The antisense nucleic acids can also be used to identify
 CC proteins used in proliferation, to express these proteins, and to obtain
 CC antibodies capable of binding to the expressed proteins. The proteins can
 CC be used to screen compounds in rational drug discovery programmes. The
 CC antisense nucleic acid sequence is also useful to screen for homologous
 CC nucleic acids which are required for cell proliferation in a wide variety
 CC of organisms. The present sequence represents an essential prokaryotic
 CC cellular proliferation protein. Note: The sequence data for this patent
 CC did not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 359 AA;
 Query Match 46.5%; Score 47; DB 4; Length 359;
 Best Local Similarity 62.5%; Pred. No. 35;
 Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 QY 1 LQRAVRLALSGFDGRV 16
 ||:|||||
 DB 126 LDRSARLSLSLGIDPRV 141
 ||:|||||
 RESULT 4
 ABU15565
 ID ABU15565 standard; protein; 359 AA.
 XX
 AC ABU15565;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Protein encoded by Prokaryotic essential gene #1092.
 XX
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX
 OS Pseudomonas aeruginosa.
 XX
 PN WO200277183-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 21-MAR-2002; 2002WO-US009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX
 DR WPI; 2003-029926/02.
 DR N-PSDB; ACA19435.
 XX
 PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 PS Claim 25; SEQ ID NO 43489; 1766pp; English.
 XX
 CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:

CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway;
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 359 AA;

Query Match 46.5%; Score 47; DB 6; Length 359;
 Best Local Similarity 62.5%; Pred. No. 35;
 Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 LQRAVRLALSGFDGRV 16
 DB 126 LQRAVRLALSGFDGRV 141

RESULT 5
 ADC94164
 ID ADC94164 standard; protein; 339 AA.

XX ADC94164;
 XX
 XX 01-JAN-2004 (first entry)
 XX
 XX E. faecium protein sequence SEQ ID 3791.
 XX
 XX Vaccine; urinary tract infection; bacteraemia; endocarditis; wound;
 KW abdominal-pelvic infection.
 XX Enterococcus faecium.
 XX US6583275-B1.
 XX
 XX 24-JUN-2003.
 XX
 XX 30-JUN-1998; 98US-00107532.
 XX
 XX 02-JUL-1997; 97US-0051571P.
 XX 14-MAY-1998; 98US-0085598P.
 XX
 XX (GENO-) GENOME THERAPEUTICS CORP.
 XX
 XX Doucette-Stamm LA, Bush D;
 XX
 XX WPI; 2003-799836/75.
 DR N-PSDB; ADC90510.
 XX
 XX New isolated nucleic acid derived from *Enterococcus faecium* encoding an
 PT Enterococcus faecium polypeptide useful for detection, prevention and

PT treatment of a pathological condition resulting from a bacterial
 PT infection.
 XX Example 1; SEQ ID NO 3791; 243pp; English.
 XX
 CC The invention relates to an isolated nucleic acid derived from
 CC *Enterococcus faecium* encoding an *Enterococcus faecium* polypeptide having
 CC one of 10 fully defined sequences given in the (or comprising 40
 CC sequential nucleotides chosen from any of the nucleic acids, its
 CC complement or sequences hybridising to it). Also included are a
 CC recombinant vector comprising the nucleic acid operably linked to a
 CC transcription regulatory element, a cell comprising the vector and a
 CC single-stranded probe comprising the nucleic acid. The nucleic acids are
 CC chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.
 CC The nucleic acids are useful for diagnosing pathological conditions
 CC resulting from *E. faecium* bacterial infection (e.g. urinary tract
 CC infection, bacteraemia, endocarditis, wounds and abdominal-pelvic
 CC infection) and for screening drugs such as agonists and antagonists. The
 CC nucleic acid is useful for recombinant production of *Candida albicans* -
 CC derived peptides or antisense polypeptides. Pharmaceutical compositions
 CC and vaccines containing the nucleic acid are useful for preventing or
 CC treating *Enterococcus faecium* infections. The present sequence represents
 CC one if the disclosed *E. faecium* proteins.
 XX
 XX Sequence 339 AA;

Query Match 45.5%; Score 46; DB 7; Length 339;
 Best Local Similarity 57.9%; Pred. No. 47;
 Matches 11; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 LQRAVRLALSGFDGRVGS 19
 DB 158 LQRTVRGLYLFSGPVGSG 176

RESULT 6
 AAG92858
 ID AAG92858 standard; protein; 229 AA.

XX AAG92858;
 XX
 XX 26-SEP-2001 (first entry)
 XX
 XX C glutamicum protein fragment SEQ ID NO: 6612.
 XX
 XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis.
 XX Corynebacterium glutamicum.
 XX EP1108790-A2.
 XX
 XX 20-JUN-2001.
 XX
 XX 18-DEC-2000; 2000EP-00127688.
 XX
 XX 16-DEC-1999; 99JP-00377484.
 XX 07-APR-2000; 2000JP-00159162.
 XX 03-AUG-2000; 2000JP-00280988.
 XX
 XX (KYOW) KYOWA HAKKO KOGYO KK.
 XX
 XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 XX
 XX WPI; 2001-376931/40.
 DR N-PSDB; AAH68077.
 XX
 XX Novel polynucleotides derived from Coryneform bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analyzing
 PT expression profile or pattern of a gene and identifying homologous gene.
 XX
 XX Claim 17; SEQ ID NO 6612; 246pp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein
 CC sequences from the Corynebacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of corynebacterium bacterium, measuring expression amount and analysing
 CC the expression profile or expression pattern of a gene derived from
 CC corynebacterium bacterium, and identifying a homologue of a gene derived from
 CC corynebacterium bacterium. Corynebacterium bacteria are useful for producing amino
 CC acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a protein described in the
 CC exemplification of the invention. Note: The sequence data for this patent
 CC did not form part of the printed specification, but was obtained in
 CC electronic format directly from the European Patent Office
 XX
 XX Sequence 229 AA;
 Query Match 45.0%; Score 45.5; DB 4; Length 229;
 Best Local Similarity 47.6%; Pred. No. 38;
 Matches 10; Conservative 4; Mismatches 4; Indels 3; Gaps 1;
 QY 2 QRAVRLALSGF---DGRVSG 19
 ||| : | : | : ||| : |
 Db 168 QRATQAFTGYLTGDLGCG 188

RESULT 7
 ABB48000
 ID ABB48000 standard; protein; 263 AA.
 XX
 AC ABB48000;
 XX
 DT 05-FEB-2002 (first entry)
 XX
 DE Listeria monocytogenes protein #704.
 XX
 KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
 KW vitamin B12; bacterial infection; disease.
 XX
 OS Listeria monocytogenes.
 XX
 PN WO200177335-A2.
 XX
 PD 18-OCT-2001.
 XX
 PF 11-APR-2001; 2001WO-FR001118.
 XX
 PR 11-APR-2000; 2000FR-00004629.
 XX
 PA (INSP) INST PASTEUR.
 XX
 PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
 PI Dussurget O, Chetoui F, Nedjari H, Glaser P, Kunet F, Cossart P;
 PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
 PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
 PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
 PI Rose M, Voss H;
 XX
 DR WPI; 2002-010914/01.
 XX
 PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
 PT and prevention of Listeria and related bacterial infections, and related
 PT polypeptides.
 XX
 PS Claim 6; SEQ ID NO 705; 192pp; French.
 XX
 CC The present invention relates to the genome sequence of Listeria
 CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
 CC it are useful for selecting probes and primers for detecting genes in L.
 CC monocytogenes and related organisms, and for studying genetic
 CC polymorphisms and other genomes. The present sequence is a protein
 CC encoded by the genome sequence of the present invention. Proteins

CC expressed from the genome sequence are useful for raising specific
 CC antibodies, identification of L. monocytogenes and related organisms, and
 CC for biosynthesis and biodegradation, especially biosynthesis of vitamin
 CC B12. The genome sequence and proteins encoded by it are also useful for
 CC selecting compounds that regulate gene expression and cell replication
 CC and modulate L. monocytogenes-related diseases. In addition, the genome
 CC sequence and proteins encoded by it are useful in pharmaceutical and
 CC vaccine compositions for the treatment or prevention of infections by L.
 CC monocytogenes and related organisms. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 263 AA;
 Query Match 44.6%; Score 45; DB 5; Length 263;
 Best Local Similarity 61.5%; Pred. No. 53;
 Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
 QY 5 VRLALSGPDGRVG 17
 :|:|:|:|:|:|
 Db 1 MRVAVSGFKGRMG 13

RESULT 8
 ABU32866
 ID ABU32866 standard; protein; 263 AA.
 XX
 AC ABU32866;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Protein encoded by Prokaryotic essential gene #18393.
 XX
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
 KW Listeria monocytogenes.
 OS
 PN WO200277183-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 21-MAR-2002; 2002WO-US009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;
 XX
 DR WPI; 2003-029926/02.
 DR N-PSDB; ACA36736.
 XX
 PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 PS Claim 25; SEQ ID NO 60790; 1766pp; English.
 XX
 CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX
 SQ Sequence 263 AA;
 Query Match 44.6%; Score 45; DB 6; Length 263;
 Best Local Similarity 61.5%; Pred. No. 53;
 Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 5 VRLALSGFDGRVG 17
 :|:|:|:|:|:|:
 Db 1 MRVAVSGFKGRMG 13

RESULT 9
 ABM67849
 ID ABM67849 standard; protein; 233 AA.

XX AC ABM67849;

XX DT 20-NOV-2003 (first entry)

XX DE Photorhabdus luminescens protein sequence #946.

XX KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
 KW detection; food; gene expression; plant; animal; microorganism; toxin;
 KW antibiotic; biopesticide; virulence factor; disease model; plague;
 KW whooping cough.

XX OS Photorhabdus luminescens.

XX PN WO200294867-A2.

XX PD 28-NOV-2002.

XX PF 07-FEB-2002; 2002WO-IB003040.

XX PR 07-FEB-2001; 2001FR-00001659.

XX (INSP) INST PASTEUR.

PA (CNRS) CNRS CENT NAT RECH SCI.

XX PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunet F, Danchin A;
 PI Buchrieser C;

XX DR WPI; 2003-148459/14.

XX PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
 PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.

XX PS Claim 2; SEQ ID NO 946; 1205pp; French.

XX CC The invention relates to the isolation of genes and their encoded

CC proteins from Photorhabdus luminescens. The isolated sequences are
 CC sources of probes and primers for detecting the genome of *P. luminescens*
 CC and related species; to study polymorphisms; for gene analysis and for
 CC detection/amplification of the genes. Antibodies (Ab) raised against the
 CC polypeptides encoded by the genes are used for detection/identification
 CC of *P. luminescens*, e.g. in foods. The genes, proteins, Ab and cells that
 CC carry a gene-containing vector are used to select compounds that
 CC modulate, regulate, induce or inhibit expression of the genes in plants,
 CC animals or microorganisms other than *P. luminescens* and are able to alter
 CC response or sensitivity to toxins and antibiotics produced by *P.*
 CC luminescens. Cells transformed to express the genes are useful for
 CC recombinant production of the proteins, particularly toxins and
 CC antibacterials useful as insecticides, bactericides and fungicides. The
 CC genes, proteins, vectors containing the genes and Ab are also useful
 CC therapeutically (to treat microbial infection by bacteria or fungi that
 CC are sensitive to *P. luminescens*-encoded toxins or antibiotics) and as
 CC biopesticides. Other uses of the genes and the proteins are as virulence
 CC factors and for identifying targets of human diseases for which *P.*
 CC luminescens is a model (particularly plague and whooping cough). This
 CC sequence represents one of the isolated *P. luminescens* proteins

XX Sequence 233 AA;

Query Match 43.6%; Score 44; DB 6; Length 233;
 Best Local Similarity 64.3%; Pred. No. 67;
 Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 RAVRLALSGFDGRV 16

Db 110 RAIRLALEGVDRRL 123
 ||:|:|:|:|:|:|:
 ||:|:|:|:|:|:|:

RESULT 10

AAU44797

ID AAU44797 standard; protein; 53 AA.

XX AC AAU44797;

XX DT 27-FEB-2002 (first entry)

XX DE Propionibacterium acnes immunogenic protein #5693.

XX KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KW dermatological; osteopathic; neuroprotectant.

XX OS Propionibacterium acnes.

XX PN WO200181581-A2.

XX PD 01-NOV-2001.

XX PF 20-APR-2001; 2001WO-US012865.

XX PR 21-APR-2000; 2000US-0199047P.

XX PR 02-JUN-2000; 2000US-0208841P.

XX PR 07-JUL-2000; 2000US-0216747P.

XX (CORI-) CORIXA CORP.

XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX WPI; 2001-616774/71.

DR N-PSDB; AAS59524.

XX PT Propionibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris.

XX PS Example 1; SEQ ID NO 5992; 1069pp; English.

CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
 CC polypeptides. The proteins and their associated DNA sequences are used in
 CC the treatment, prevention and diagnosis of medical conditions caused by
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
 CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
 CC P. acnes is also involved in infections of bone, joints and the central
 CC nervous system, however it is particularly involved in the inflammatory
 CC lesions associated with acne vulgaris. A method for detecting the
 CC presence or absence of P. acnes in a patient comprises contacting a
 CC sample with a binding agent that binds to the proteins of the invention
 CC and determining the amount of bound protein in the sample. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC specific for P. acnes proteins. These antibodies can be used to
 CC downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
 CC this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 53 AA;

Query Match 43.1%; Score 43.5; DB 4; Length 53;
 Best Local Similarity 61.1%; Pred. No. 17;
 Matches 11; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

QY 4 AVRLLSGFDG-RVSGGC 20
 :||:|||||
 Db 8 SVRVALSTEDGVRVSMRC 25

RESULT 11
 ABM41316
 ID ABM41316 standard; protein; 53 AA.

AC ABM41316;

DT 20-OCT-2003 (first entry)

DE Propionibacterium acnes predicted ORF-encoded polypeptide #5992.

XX Acne vulgaris; antisborrheic; dermatological; antibacterial;

KW immunostimulant; immune response; vaccine.

XX Propionibacterium acnes.

XX WO2003033515-A1.

XX 24-APR-2003.

XX 11-OCT-2002; 2002WO-US032727.

XX 15-OCT-2001; 2001US-00978825.

XX (CORI-) CORIXA CORP.

XX Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;

PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;

PI Barth B, Valliave-Douglass J;

XX WPI; 2003-381789/36.

DR N-PSDB; ACP64453.

XX New Propionibacterium acnes polypeptides and polynucleotides encoding the
 PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
 PT or for stimulating an immune response specific for a P. acnes protein.

XX Example 1; SEQ ID NO 5992; 1481pp; English.

XX The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
 CC encoding a Propionibacterium acnes protein. The invention also relates to
 CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to

CC immunogenic fragments of P. acnes polypeptides. The invention
 CC additionally encompasses expression vectors and host cells comprising a
 CC polynucleotide of the invention; antibodies against polypeptides of the
 CC invention; fusion proteins comprising a polypeptide of the invention; a
 CC method for stimulating an immune response specific for a P. acnes
 CC polypeptide and an isolated T cell population comprising T cells prepared
 CC via this method; a vaccine composition (comprising P. acnes polypeptides,
 CC polynucleotides, antibodies, fusion proteins, T cell populations, or
 CC antigen-presenting cells that express the polypeptide); a method and kit
 CC for detecting or determining the presence or absence of P. acnes in a
 CC patient; and a method for inhibiting the development of P. acnes in a
 CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
 CC proteins, T cell populations or antigen-presenting cells that express the
 CC polypeptides are useful for diagnosing, preventing or treating acne
 CC vulgaris, or for stimulating an immune response specific for a P. acnes
 CC protein. The polynucleotides can also be used as probes or primers for
 CC nucleic acid hybridization. The vaccine composition is useful for the
 CC stimulation of an immune response against P. acnes, or for treating acne,
 CC and the kit is useful for performing a diagnostic assay. The present
 CC sequence represents a polypeptide predicted to be encoded by an ORF (open
 CC reading frame) contained within the P. acnes polynucleotides of the
 CC invention. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 53 AA;

Query Match 43.1%; Score 43.5; DB 6; Length 53;
 Best Local Similarity 61.1%; Pred. No. 17;
 Matches 11; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

QY 4 AVRLLSGFDG-RVSGGC 20
 :||:|||||
 Db 8 SVRVALSTEDGVRVSMRC 25

RESULT 12

AAU63931

ID AAU63931 standard; protein; 141 AA.

AC AAU63931;

DT 27-FEB-2002 (first entry)

DE Propionibacterium acnes immunogenic protein #24827.

KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KW dermatological; osteopathic; neuroprotectant.

OS Propionibacterium acnes.

XX WO200181581-A2.

XX 01-NOV-2001.

XX 20-APR-2001; 2001WO-US012865.

XX 21-APR-2000; 2000US-0199047P.

XX 02-JUN-2000; 2000US-0208841P.

XX 07-JUL-2000; 2000US-0216747P.

XX (CORI-) CORIXA CORP.

XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX WPI; 2001-616774/71.

DR N-PSDB; AAS59637.

XX Propionibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for

PT treating acne vulgaris.

PS Example 1; SEQ ID NO 25126; 1069pp; English.

XX

CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic

CC polypeptides. The proteins and their associated DNA sequences are used in

CC the treatment, prevention and diagnosis of medical conditions caused by

CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,

CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.

CC P. acnes is also involved in infections of bone, joints and the central

CC nervous system, however it is particularly involved in the inflammatory

CC lesions associated with acne vulgaris. A method for detecting the

CC presence or absence of P. acnes in a patient comprises contacting a

CC sample with a binding agent that binds to the proteins of the invention

CC and determining the amount of bound protein in the sample. The

CC polypeptides may be used as antigens in the production of antibodies

CC specific for P. acnes proteins. These antibodies can be used to

CC downregulate expression and activity of P. acnes polypeptides and

CC therefore treat P. acnes infections. The antibodies may also be used as

CC diagnostic agents for determining P. acnes presence, for example, by

CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for

CC this patent did not form part of the printed specification, but was

CC obtained in electronic format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences

XX

XX Sequence 141 AA;

XX

Query Match 42.6%; Score 43; DB 4; Length 141;

Best Local Similarity 55.6%; Pred. No. 57;

Matches 10; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 LQRAVRLALSGDFGRVGS 18

DB 28 VQHRVRGFLGLDGMVGS 45

RESULT 13

ABM60450

ID ABM60450 standard; protein; 141 AA.

XX

AC ABM60450;

XX

DT 20-OCT-2003 (first entry)

XX

DE Propionibacterium acnes predicted ORF-encoded polypeptide #25126.

XX

KW Acne vulgaris; antisborrheic; dermatological; antibacterial;

KW immunostimulant; immune response; vaccine.

XX

OS Propionibacterium acnes.

XX

PN WO2003033515-A1.

XX

PD 24-APR-2003.

XX

PF 11-OCT-2002; 2002WO-US032727.

XX

PR 15-OCT-2001; 2001US-00978825.

XX

XX (CORI-) CORIXA CORP.

PA Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;

PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;

PI Barth B, Vallieve-Douglas J;

XX

XX WPI; 2003-381789/36.

DR N-PSDB; ACF64566.

XX

XX New Propionibacterium acnes polypeptides and polynucleotides encoding the

PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,

PT or for stimulating an immune response specific for a P. acnes protein.

XX

XX Example 1; SEQ ID NO 25126; 1481pp; English.

PS

XX

XX The invention relates to an isolated polynucleotide (ACF64435-ACF64733)

CC encoding a Propionibacterium acnes protein. The invention also relates to

CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to

CC immunogenic fragments of P. acnes polypeptides. The invention

CC additionally encompasses expression vectors and host cells comprising a

CC polynucleotide of the invention; antibodies against polypeptides of the

CC invention; fusion proteins comprising a polypeptide of the invention; a

CC method for stimulating an immune response specific for a P. acnes

CC polypeptide and an isolated T cell population comprising T cells prepared

CC via this method; a vaccine composition comprising P. acnes polypeptides,

CC polynucleotides, antibodies, fusion proteins, T cell populations, or

CC antigen-presenting cells that express the polypeptide; a method and kit

CC for detecting or determining the presence or absence of P. acnes in a

CC patient; and a method for inhibiting the development of P. acnes in a

CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion

CC proteins, T cell populations or antigen-presenting cells that express the

CC polypeptides are useful for diagnosing, preventing or treating acne

CC vulgaris, or for stimulating an immune response specific for a P. acnes

CC protein. The polynucleotides can also be used as probes or primers for

CC nucleic acid hybridisation. The vaccine composition is useful for the

CC stimulation of an immune response against P. acnes, or for treating acne,

CC and the kit is useful for performing a diagnostic assay. The present

CC sequence represents a polypeptide predicted to be encoded by an ORF (open

CC reading frame) contained within the P. acnes polynucleotides of the

CC invention. Note: The sequence data for this patent did not form part of

CC the printed specification, but was obtained in electronic format directly

CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX

XX Sequence 141 AA;

XX

Query Match 42.6%; Score 43; DB 6; Length 141;

Best Local Similarity 55.6%; Pred. No. 57;

Matches 10; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 LQRAVRLALSGDFGRVGS 18

DB 28 VQHRVRGFLGLDGMVGS 45

RESULT 14

AAR98742

ID AAR98742 standard; protein; 228 AA.

XX

AC AAR98742;

XX

DT 09-DEC-1996 (first entry)

XX

DE Tazarotene induced gene 1 (TIG1) encoded protein.

XX

KW Tazarotene induced gene 1; TIG1; retinoid; psoriasis; antiisporiatic;

KW fibroblast; keratinocyte; skin raft culture; retinoic acid receptor.

XX

OS Homo sapiens.

XX

EH Key Location/Qualifiers

FT Domain 16..34

FT /label= Transmembrane domain

FT /note= "putative transmembrane domain"

FT Modified-site 142..144

FT /label= Glycosylation

FT /note= "N-linked glycosylation site"

XX

PN WO9623080-A1.

XX

XX 01-AUG-1996.

PD

XX

XX 26-JAN-1996; 96WO-US001339.

PF

XX

PR 27-JAN-1995; 95US-00379280.

XX

XX (ALLR) ALLERGAN.

XX

PI Nagpal S, Chandraratna RA;
 XX WPI; 1996-362702/36.
 DR N-PSDB; AAT38115.
 XX Retinoid-inducible polynucleotide, designated Tazarotene Induced Gene
 PT (TGII) - having low basal expression, used in assay systems as an
 PT indicator of retinoid action in psoriatic skin cultures.
 PS Example 7; Page 40-41; 50pp; English.
 XX A putative transmembrane protein (AAR98742) is encoded by novel human
 CC TIG1 (tazarotene induced gene 1) cDNA (AAT38115). TIG1 mRNA is strongly
 CC induced from a low basal level upon treatment of skin raft cultures with
 CC the retinoic acid receptor (RAR) beta/gamma selective antipsoriatic
 CC synthetic retinoid AGN-190168 (ethyl 6-(2-(4,4) dimethyl-thiochroman-6-
 CC yl) ethynyl-nicotinate). Inducible expression of TIG1 mRNA in psoriatic
 CC skin raft cultures has also been demonstrated. The low basal expression
 CC of TIG1 is partic. advantageous when used as an indicator of retinoid
 CC action in psoriatic skin cultures
 XX Sequence 228 AA;
 SQ

Query Match 42.6%; Score 43; DB 2; Length 228;
 Best Local Similarity 61.1%; Pred. No. 96;
 Matches 11; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 QY 1 LQAVRLALSGPGRVGS 18
 ||||| ||| | : |||
 DB 59 LQAVRLALHFFNFRSGS 76

RESULT 15
 ABU48621
 ID ABU48621 standard; protein; 287 AA.
 XX
 AC ABU48621;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Protein encoded by Prokaryotic essential gene #34148.
 XX
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
 XX
 OS Treponema pallidum.
 XX
 PN WO20027183-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 21-MAR-2002; 2002WO-US009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KU, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX
 DR WPI; 2003-029926/02.
 DR N-PSDB; ACAS2491.
 XX
 XX New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 PS Claim 25; SEQ ID NO 76545; 1766pp; English.
 XX
 CC The invention relates to an isolated nucleic acid comprising any one of

CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
 CC the target prokaryotic essential genes. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 287 AA;
 QY Query Match 42.6%; Score 43; DB 6; Length 287;
 Best Local Similarity 58.3%; Pred. No. 1-2e+02;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 DB 9 LSGFDRGVSGC 20
 : || ||| : ||
 193 IGGFAGSVGAGC 204

RESULT 16
 ABB48949
 ID ABB48949 standard; protein; 352 AA.
 XX
 AC ABB48949;
 XX
 DT 05-FEB-2002 (first entry)
 XX
 DE *Listeria monocytogenes* protein #1653.
 XX
 KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
 XX vitamin B12; bacterial infection; disease.
 XX
 OS *Listeria monocytogenes*.
 XX
 PN WO200177335-A2.
 XX
 PD 18-OCT-2001.
 XX
 PF 11-APR-2001; 2001WO-FR001118.
 XX
 PR 11-APR-2000; 2000FR-00004629.
 XX
 PA (INSP) INST PASTEUR.
 XX
 XX Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
 PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunet F, Cossart P;
 PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
 PI Dominguez-Bernal G, Garrido-Garcia P, Pierrez-Martinez A, Amend A;
 PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;

PR 14-AUG-2000; 2000US-0225758P.
 PR 14-AUG-2000; 2000US-0225759P.
 PR 18-AUG-2000; 2000US-0226279P.
 PR 22-AUG-2000; 2000US-0226681P.
 PR 22-AUG-2000; 2000US-0226988P.
 PR 22-AUG-2000; 2000US-0227182P.
 PR 23-AUG-2000; 2000US-0227009P.
 PR 30-AUG-2000; 2000US-0228924P.
 PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 05-SEP-2000; 2000US-0229513P.
 PR 06-SEP-2000; 2000US-0230437P.
 PR 06-SEP-2000; 2000US-0230438P.
 PR 08-SEP-2000; 2000US-0231242P.
 PR 08-SEP-2000; 2000US-0231243P.
 PR 08-SEP-2000; 2000US-0231244P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 08-SEP-2000; 2000US-0231414P.
 PR 08-SEP-2000; 2000US-0232080P.
 PR 08-SEP-2000; 2000US-0232081P.
 PR 12-SEP-2000; 2000US-0231968P.
 PR 14-SEP-2000; 2000US-0232397P.
 PR 14-SEP-2000; 2000US-0232398P.
 PR 14-SEP-2000; 2000US-0232399P.
 PR 14-SEP-2000; 2000US-0232400P.
 PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0232402P.
 PR 14-SEP-2000; 2000US-0232403P.
 PR 14-SEP-2000; 2000US-0232404P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 25-SEP-2000; 2000US-0234998P.
 PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 27-SEP-2000; 2000US-0235836P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239335P.
 PR 13-OCT-2000; 2000US-0239337P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-0246179P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.

PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249264P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0251989P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM;
 DR WPI; 2001-502630/55.
 DR N-PSDB; AAK88256.
 XX
 PT Polynucleotides encoding digestive system antigens, useful for
 PT diagnosing, treating, preventing and/or prognosing disorders of the
 PT digestive system, particularly cancer and cancer metastases.
 XX
 PS Claim 11; SEQ ID NO 1832; 986pp; English.
 XX
 CC The present invention provides the protein and coding sequences of a
 CC number of human digestive system antigens. These can be used in the
 CC diagnosis, treatment and prevention of digestive system disorders,
 CC including cancer, Meckel's diverticulum, bacterial or parasitic
 CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
 CC ulcerative colitis. The present sequence is a digestive system antigen of
 CC the invention
 XX
 SQ Sequence 78 AA;
 Query Match 42.1%; Score 42.5; DB 4; Length 78;
 Best Local Similarity 52.2%; Pred. No. 37;
 Matches 12; Conservative 1; Mismatches 7; Indels 3; Gaps 1;
 QY 1 LQRAVRLAL---SGFDGRVGSQC 20
 ||| ||| ||| ||| : |||
 Db 36 LQSQVSLALCHPSSVDGWLGSQC 58
 RESULT 21
 AAU22528
 ID AAU22528 standard; protein; 78 AA.
 XX
 AC AAU22528;

XX 17-DEC-2001 (first entry)
DT Novel human colon associated polypeptide #61.
XX
DE
DE Human; colon cancer; congenital abnormality; infection; colitis;
KW inflammatory bowel disease; IBD; neoplastic disorder; gene therapy;
KW intestinal inflammatory disorder; malabsorption syndrome; gastric;
KW sigmoid disease; antibacterial; antiviral; antiinflammatory; cytostatic.
XX
OS Homo sapiens.
XX
PN WO20015302-A2.
XX
XX 02-AUG-2001.
PD
XX 17-JAN-2001; 2001WO-US001240.
PF
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-MAR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225467P.
PR 14-AUG-2000; 2000US-0225468P.
PR 14-AUG-2000; 2000US-022570P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226686P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231444P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0234848P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235835P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.

```
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465567/50.
DR N-PSDB; AAS39408.
XX
XX Isolated polypeptide for treating, preventing and/ or prognosing
PT disorders related to the colon including colon cancers and also for
PT testing and detection e.g. diagnosis.
XX
XX Claim 11; SEQ ID NO 305; 562pp; English.
XX
XX The present invention relates to the isolation of novel human colon
CC associated polypeptides, and the cDNA (AAS39348-AAS39581) and genomic
CC sequences encoding for them. The sequences of the invention are useful in
CC the diagnosis, treatment, prevention and/or prognosis of disorders of the
CC colon including colon cancer, congenital abnormalities (e.g. atresia and
CC stenosis), bacterial and viral infections, inflammatory bowel disease
CC (IBD), neoplastic cell disorders (e.g. polyps and adenomas, intestinal
CC inflammatory disorders, colitis, colonic inflammation, diarrhoea and
CC dysentery, malabsorption syndromes (e.g. lactose intolerance), intestinal
CC obstruction and sigmoid diseases. The polynucleotide sequences of the
CC invention can also be used in gene therapy. AAU22468-AAU22701 represent
CC the novel human colon associated polypeptides of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 78 AA;
XX
XX Query Match 42.1%; Score 42.5; DB 4; Length 78;
XX Best Local Similarity 52.2%; Pred. No. 37;
XX Matches 12; Conservative 1; Mismatches 7; Indels 3; Gaps 1;
XX
XX 1 LQRAVRLAL---SGFGRVCSGC 20
XX 36 LQSQVSLALCHPSSVDGLGSSC 58
XX
XX RESULT 22
XX ADB32368
XX ID ADB32368 standard; protein; 78 AA.
XX AC ADB32368;
XX XX
XX 04-DEC-2003 (first entry)
XX DE
XX Human novel colon related polypeptide SEQ ID NO 305.
XX
XX gene therapy; cancer; liver disorder; hepatitis; neural disorder;
XX Alzheimer's disease; human; colon.
XX Homo sapiens.
XX
XX US2003050231-A1.
XX
XX 13-MAR-2003.
XX
XX 17-JAN-2001; 2001US-00764872.
XX
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184664P.
XX
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216800P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226688P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0234080P.
PR 08-SEP-2000; 2000US-0234081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0234010P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
```


PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 01-NOV-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251198P.
PR 06-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 11-DEC-2000; 2000US-0251990P.
PR 05-JAN-2001; 2001US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX (ROSE/) ROSEN C A.
PA (RUBE/) RUBEN S M.
PA (BARA/) BARASH S C.
XX Rosen CA, Ruben SM, Barash SC;
PI
XX WPI; 2003-625420/59.
DR N-PSDB; ADB32134.
XX
XX New nucleic acid molecule, useful for preparing a medicament for
PT preventing, treating or ameliorating a medical condition e.g. cancer,
PT liver disorders or neural disorders.

XX Claim 12; SEQ ID NO 305; 216pp; English.
PS
XX The invention relates to an isolated nucleic acid molecule encoding a
CC polypeptide. The nucleic acid is useful for preparing a medicament for
CC preventing, treating or ameliorating a medical condition e.g. cancer,
CC liver disorders such as hepatitis or neural disorders such as Alzheimer's
CC disease. The present sequence represents the amino acid sequence of a
CC human novel colon related polypeptide. Note: The sequence data for this
CC patent did not form part of the printed specification but was obtained in
CC electronic format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=20030050231.
XX
SQ Sequence 78 AA;
Query Match 42.1%; Score 42.5; DB 7; Length 78;
Best Local Similarity 52.2%; Pred. No. 37;
Matches 12; Conservative 1; Mismatches 7; Indels 3; Gaps 1;
QY 1 LQRAVRLAL---SGFDGRVSGSC 20
Db 36 LQSQVSLALCHPSSVDGMLGSSC 58
RESULT 23
AAM00762
ID AAM00762 standard; peptide; 17 AA.
XX
AC AAM00762;
XX
DT 01-OCT-2001 (first entry)
XX Human bone marrow protein, SEQ ID NO: 125.
XX Human; bone marrow; antiinflammatory; cytostatic; neuroprotective;
KW antiviral; antibacterial; antifungal; anti-HIV; haemostatic;
KW immunosuppressive; gene therapy; cytokine cell proliferation;
KW cell differentiation modulator; immune disorder; infection; cancer;
KW human immunodeficiency virus; HIV; autoimmune disorder; haemophilia.
OS Homo sapiens.
XX
PN WO200153453-A2.
XX
PD 26-JUL-2001.
XX
PF 23-DEC-2000; 2000WO-US034960.
XX
PR 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 30-NOV-2000; 2000US-0250583P.
XX (HYSE-) HYSEQ INC.
XX
PI Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Chen R, Ma Y;
PI Ren F, Wang J, Werhman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Drmanac RT;
XX
DR WPI; 2001-488707/53.
DR N-PSDB; AAH99881.
XX
PT Novel bone-marrow-expressed polynucleotides and polypeptides, useful for
PT treating e.g. cancer and immune deficiency disorders.
XX
PS Claim 10; Page 296; 648pp; English.
XX The present sequence is one of 251 novel human polypeptides encoded by a
CC

CC bone marrow-expressed polynucleotide. The polynucleotide and the
 CC polypeptide encoded by it are useful in the treatment of various immune
 CC deficiencies and disorders. The deficiencies and disorders may be
 CC genetic, may be caused by a viral (e.g. HIV), bacterial or fungal
 CC infection, or may result from an autoimmune disorder, a coagulation
 CC disorder (e.g. haemophilia), inhibition of tumour cell proliferation,
 CC suppression of an inflammatory response or treatment of a nervous system
 CC disorder such as Alzheimer's disease. Detection of the presence or
 CC increased expression of the polynucleotide or the protein it encodes is
 CC useful for the diagnosis and/or prognosis of one or more types of cancer.
 CC The polynucleotide and polypeptide can be used as nutritional sources or
 CC supplements and in the screening of chemical compounds as potential drugs
 XX
 SQ Sequence 17 AA;

Query Match 41.6%; Score 42; DB 4; Length 17;
 Best Local Similarity 57.1%; Pred. No. 8.6;
 Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 6 RLALSGFGDGRVSG 19
 | : : | : : | : : |
 Db 3 RWMAGEDGVGCG 16

RESULT 24
 AAU44009
 ID AAU44009 standard; protein; 133 AA.

XX AC AAU44009;
 XX DT 27-FEB-2002 (first entry)
 XX DE Propionibacterium acnes immunogenic protein #4905.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KW dermatological; osteopathic; neuroprotectant.
 XX OS Propionibacterium acnes.

XX PN WO200181581-A2.

XX PD 01-NOV-2001.

XX PF 20-APR-2001; 2001WO-US012865.

XX PR 21-APR-2000; 2000US-0199047P.

XX PR 02-JUN-2000; 2000US-0208841P.

XX PR 07-JUL-2000; 2000US-0216747P.

XX PA (CORI-) CORIXA CORP.

XX PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX DR WPI; 2001-616774/71.

XX DR N-PSDB; AAS59521.

XX PT Propionibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris.

XX PS Example 1; SEQ ID NO 5204; 1069pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
 CC polypeptides. The proteins and their associated DNA sequences are used in
 CC the treatment, prevention and diagnosis of medical conditions caused by
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
 CC P. acnes is also involved in infections of bone, joints and the central
 CC nervous system, however it is particularly involved in the inflammatory
 CC lesions associated with acne vulgaris. A method for detecting the

CC presence or absence of P. acnes in a patient comprises contacting a
 CC sample with a binding agent that binds to the proteins of the invention
 CC and determining the amount of bound protein in the sample. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC specific for P. acnes proteins. These antibodies can be used to
 CC downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
 CC this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 133 AA;

Query Match 41.6%; Score 42; DB 4; Length 133;
 Best Local Similarity 64.3%; Pred. No. 78;
 Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 3 RAVRLALSGFDGRV 16
 | | | | | : | : | : |
 Db 110 RATRLTLDDFDGAV 123

RESULT 25
 ABM40528
 ID ABM40528 standard; protein; 133 AA.

XX AC ABM40528;

XX DT 20-OCT-2003 (first entry)

XX DE Propionibacterium acnes predicted ORF-encoded polypeptide #5204.

XX Acne vulgaris; antisborrheic; dermatological; antibacterial;
 KW immunostimulant; immune response; vaccine.

XX OS Propionibacterium acnes.

XX PN WO2003033515-A1.

XX PD 24-APR-2003.

XX PF 11-OCT-2002; 2002WO-US032727.

XX PR 15-OCT-2001; 2001US-00978825.

XX PA (CORI-) CORIXA CORP.

XX PI Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;
 PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;
 PI Barth B, Vallieue-Douglas J;

XX DR WPI; 2003-381789/36.

XX DR N-PSDB; ACF64450.

XX PT New Propionibacterium acnes polypeptides and polynucleotides encoding the
 PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
 PT or for stimulating an immune response specific for a P. acnes protein.

XX PS Example 1; SEQ ID NO 5204; 1481pp; English.

XX The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
 CC encoding a Propionibacterium acnes protein. The invention also relates to
 CC polypeptides encoded by the polynucleotides (ABM35624-ABM4536) and to
 CC immunogenic fragments of P. acnes polypeptides. The invention
 CC additionally encompasses expression vectors and host cells comprising a
 CC polynucleotide of the invention; antibodies against polypeptides of the
 CC invention; fusion proteins comprising a polypeptide of the invention; a
 CC method for stimulating an immune response specific for a P. acnes
 CC polypeptide and an isolated T cell population comprising T cells prepared
 CC via this method; a vaccine composition (comprising P. acnes polypeptides,
 CC polynucleotides, antibodies, fusion proteins, T cell populations, or

XX Herbicidally active polypeptide SEQ ID NO 1821.
DE
XX
XX KW Herbicidal; plant; agriculture; herbicide.
XX
XX OS Arabidopsis thaliana.
XX
XX PN WO200210210-A2.
XX
XX PD 07-FEB-2002.
XX
XX 28-AUG-2001; 2001WO-EP009892.
XX
XX PR 28-AUG-2001; 2001WO-EP009892.
XX
XX PA (FARB) BAYER AG.
XX
XX Tietjen K, Weidler M;
XX
XX WPI; 2002-269010/31.
XX
XX Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms.
XX
XX Claim 5; SEQ ID NO 1821; 261pp + Sequence Listing; English.
XX
XX The invention relates to identifying target proteins (ABB90790-ABB94016)
CC for herbicidally active compounds, comprising aligning and comparing
CC nucleic acid or amino acid sequences from plant with nucleic acid or
CC amino acid sequences from non-plant organisms using suitable search
CC parameters, where plant sequences having an E-value greater by a factor
CC of 3 than the E-value of most similar non-plant sequences are selected.
CC The polypeptides or nucleic acids encoding them are useful for
CC identifying modulators. The identified modulators are useful as
CC herbicides
XX
XX SQ Sequence 466 AA;
Query Match 41.6%; Score 42; DB 5; Length 466;
Best Local Similarity 56.2%; Pred. No. 3e+02; Mismatches 5; Indels 0; Gaps 0;
Matches 9; Conservative 2;
QY 4 AVRLLSGFDGRVSG 19
DB 252 AVNAVVGFDGLVFTG 267

Search completed: May 13, 2004, 06:52:35
Job time : 52 secs

ALIGNMENTS

RESULT 1
US-08-374-560-7
; Sequence 7, Application US/08374560
; Patent No. 5882645
; GENERAL INFORMATION:
; APPLICANT: TOTH, Istvan
; APPLICANT: GIBBONS, William Anthony
; TITLE OF INVENTION: PEPTIDE COMPOUNDS
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen
; STREET: 1180 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10036-8403
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/374,560
; FILING DATE: 13-Mar-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9215780.9
; FILING DATE: 24-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Meilman, Edward A.
; REGISTRATION NUMBER: 24,735
; REFERENCE/DOCKET NUMBER: P/365-302
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 382-0700
; TELEFAX: (212) 382-0888
; TELEX: 236925
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
US-08-374-560-7

Query Match 75.2%; Score 76; DB 2; Length 19;
Best Local Similarity 94.1%; Pred. No. 1.9e-06;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 RAVRLALSGDGRVSG 19
Db 3 RAVRLALSGDGRVSG 19

RESULT 2
US-09-252-991A-31177
; Sequence 31177, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31177
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31177

Query Match 46.5%; Score 47; DB 4; Length 432;
Best Local Similarity 62.5%; Pred. No. 6.5;
Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 LQRAVRLALSGDGRV 16
Db 199 LDRSARLSLSDPRV 214

RESULT 3
US-09-107-532A-3791
; Sequence 3791, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3791:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 339 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...339
; SEQUENCE DESCRIPTION: SEQ ID NO: 3791:
US-09-107-532A-3791

Query Match 45.5%; Score 46; DB 4; Length 339;
Best Local Similarity 57.9%; Pred. No. 7.3;
Matches 11; Conservative 1; Mismatches 7; Indels 0; Gaps 0;


```
/ COUNTRY: U.S.A.
/ ZIP: 92660
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: DOS
/ SOFTWARE: FastSeq Version 1.5
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/379,280
/ FILING DATE:
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER:
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Altman, Daniel E
/ REGISTRATION NUMBER: 34,115
/ REFERENCE/DOCKET NUMBER: ALRGN.057A
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 714-760-0404
/ TELEFAX: 714-760-9502
/ TELEX:
/ INFORMATION FOR SEQ ID NO: 5:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 228 amino acids
/ TYPE: amino acids
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
/ FRAGMENT TYPE: internal
/ ORIGINAL SOURCE:
/ US-08-379-280-5

Query Match 42.6%; Score 43; DB 1; Length 228;
Best Local Similarity 61.1%; Pred. No. 15;
Matches 11; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 LQRAVRLALSGFGRVGS 18
||:|||||:|:|
DB 59 LQAVRAALHFFNRSGS 76

RESULT 9
US-09-252-991A-29999
; Sequence 29999, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 29999
; LENGTH: 274
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29999

Query Match 42.6%; Score 43; DB 4; Length 374;
Best Local Similarity 44.0%; Pred. No. 26;
Matches 11; Conservative 3; Mismatches 5; Indels 6; Gaps 1;

QY 1 LQRAVRLALSGFD-----GRVSG 19
|:|:|:|:|
DB 156 LEGAAGLAITAGLDGFALEGRVGG 180

/ COUNTRY: U.S.A.
/ ZIP: 92660
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: DOS
/ SOFTWARE: FastSeq Version 1.5
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/379,280
/ FILING DATE:
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER:
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Altman, Daniel E
/ REGISTRATION NUMBER: 34,115
/ REFERENCE/DOCKET NUMBER: ALRGN.057A
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 714-760-0404
/ TELEFAX: 714-760-9502
/ TELEX:
/ INFORMATION FOR SEQ ID NO: 5:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 228 amino acids
/ TYPE: amino acids
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
/ FRAGMENT TYPE: internal
/ ORIGINAL SOURCE:
/ US-08-379-280-5

Query Match 42.6%; Score 43; DB 4; Length 402;
Best Local Similarity 64.3%; Pred. No. 29;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 VRLALSGFGRVGS 18
||:|:|:|:|
DB 366 VRLLAGIDRRVGA 379

RESULT 11
US-09-489-039A-9982
; Sequence 982, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 9982
; LENGTH: 435
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-9982

Query Match 42.6%; Score 43; DB 4; Length 435;
Best Local Similarity 50.0%; Pred. No. 32;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 QRAVRLALSGFGRVSG 19
||:|:|:|:|
DB 281 QRVGHHLLGGFKRLGNG 298

RESULT 12
US-09-252-991A-24289
; Sequence 24289, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
```



```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 607 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-537-361E-10

Query Match 41.6%; Score 42; DB 3; Length 607;
Best Local Similarity 60.0%; Pred. No. 69;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 5 VRLALSGFDGRVGS 19
Db 126 VQGLSLGYGGRGSG 140

RESULT 17
US-08-990-470A-4
; Sequence 4, Application US/08990470A
; Patent No. 6123942
; GENERAL INFORMATION:
; APPLICANT: Stojiljkovic, Igor
; APPLICANT: So, Magdalene
; APPLICANT: Hwa, Vivian
; APPLICANT: Heffron, Fred
; APPLICANT: Nassif, Xavier
; TITLE OF INVENTION: No. 6123942el Bacterial Hemoglobin Receptor
; TITLE OF INVENTION: Genes and Uses
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive, 32nd Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/990,470A
; FILING DATE: 15-DEC-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6123942nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 94,784-K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 607 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-990-470A-4

Query Match 41.6%; Score 42; DB 3; Length 607;
Best Local Similarity 60.0%; Pred. No. 69;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 5 VRLALSGFDGRVGS 19
Db 126 VQGLSLGYGGRGSG 140

RESULT 18
US-08-817-707-10
; Sequence 10, Application US/08817707
; Patent No. 6277382
```

```
; GENERAL INFORMATION:
; APPLICANT: Stojiljkovic, Igor
; APPLICANT: So, Magdalene
; APPLICANT: Hwa, Vivian
; APPLICANT: Heffron, Fred
; APPLICANT: Nassif, Xavier
; TITLE OF INVENTION: No. 6277382el Bacterial Hemoglobin Receptor
; TITLE OF INVENTION: Genes and Uses
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive, 32nd Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,707
; FILING DATE: 19-AUG-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6277382nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 94,784-J
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 607 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-817-707-10

Query Match 41.6%; Score 42; DB 3; Length 607;
Best Local Similarity 60.0%; Pred. No. 69;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 5 VRLALSGFDGRVGS 19
Db 126 VQGLSLGYGGRGSG 140

RESULT 19
US-09-074-658-75
; Sequence 75, Application US/09074658
; Patent No. 6184371
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M
; APPLICANT: Run-Pan Du
; APPLICANT: Quijun Wang
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA
; NUMBER OF SEQUENCES: 78
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
```

```
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/074,658
/ FILING DATE: 08-MAY-1998
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Stewart, Michael I
/ REGISTRATION NUMBER: 24,973
/ REFERENCE/DOCKET NUMBER: 1038-795
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (416) 595-1155
/ TELEFAX: (416) 595-1163
/ INFORMATION FOR SEQ ID NO: 75:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 941 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/
/ US-09-074-658-75
/
/ Query Match 41.6%; Score 42; DB 3; Length 941;
/ Best Local Similarity 60.0%; Pred. No. 1.2e+02;
/ Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
/
/ QY 5 VRLALSGFDGRVGS 19
/ Db 122 VQSLSGYGGGSG 136
/
/ RESULT 20
/ US-08-867-941-23
/ Sequence 23, Application US/08867941
/ Patent No. 5977337
/ GENERAL INFORMATION:
/ APPLICANT: Loosmore, Sheena M
/ APPLICANT: Du, Run-Pan
/ APPLICANT: Wang, Quijun
/ APPLICANT: Yang, Yan-Ping
/ APPLICANT: Klein, Michel H
/ TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA
/ NUMBER OF SEQUENCES: 67
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Sim & McBurney
/ STREET: 6th Floor, 330 University Avenue
/ CITY: Toronto
/ STATE: Ontario
/ COUNTRY: Canada
/ ZIP: M5G 1R7
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/867,941
/ FILING DATE: 03-JUN-1997
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Stewart, Michael I
/ REGISTRATION NUMBER: 24,973
/ REFERENCE/DOCKET NUMBER: 1038-681 MIS:jb
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (416) 595-1155
/ TELEFAX: (416) 595-1163
/ INFORMATION FOR SEQ ID NO: 24:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 944 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/
/ US-08-867-941-24
/
/ Query Match 41.6%; Score 42; DB 2; Length 944;
/ Best Local Similarity 60.0%; Pred. No. 1.2e+02;
/ Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
/
/ QY 5 VRLALSGFDGRVGS 19
/ Db 122 VQSLSGYGGGSG 136
/
/ RESULT 20
/ US-08-867-941-23
/ Sequence 23, Application US/08867941
/ Patent No. 5977337
/ GENERAL INFORMATION:
/ APPLICANT: Loosmore, Sheena M
/ APPLICANT: Du, Run-Pan
/ APPLICANT: Wang, Quijun
/ APPLICANT: Yang, Yan-Ping
/ APPLICANT: Klein, Michel H
/ TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA
/ NUMBER OF SEQUENCES: 67
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Sim & McBurney
/ STREET: 6th Floor, 330 University Avenue
/ CITY: Toronto
/ STATE: Ontario
/ COUNTRY: Canada
/ ZIP: M5G 1R7
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/867,941
/ FILING DATE: 03-JUN-1997
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Stewart, Michael I
/ REGISTRATION NUMBER: 24,973
/ REFERENCE/DOCKET NUMBER: 1038-681 MIS:jb
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (416) 595-1155
/ TELEFAX: (416) 595-1163
/ INFORMATION FOR SEQ ID NO: 23:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 944 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/
/ US-08-867-941-23
/
/ Query Match 41.6%; Score 42; DB 2; Length 944;
/ Best Local Similarity 60.0%; Pred. No. 1.2e+02;
```

```
/ Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
/
/ QY 5 VRLALSGFDGRVGS 19
/ Db 127 VQSLSGYGGGSG 141
/
/ RESULT 21
/ US-08-867-941-24
/ Sequence 24, Application US/08867941
/ Patent No. 5977337
/ GENERAL INFORMATION:
/ APPLICANT: Loosmore, Sheena M
/ APPLICANT: Du, Run-Pan
/ APPLICANT: Wang, Quijun
/ APPLICANT: Yang, Yan-Ping
/ APPLICANT: Klein, Michel H
/ TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA
/ NUMBER OF SEQUENCES: 67
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Sim & McBurney
/ STREET: 6th Floor, 330 University Avenue
/ CITY: Toronto
/ STATE: Ontario
/ COUNTRY: Canada
/ ZIP: M5G 1R7
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/867,941
/ FILING DATE: 03-JUN-1997
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Stewart, Michael I
/ REGISTRATION NUMBER: 24,973
/ REFERENCE/DOCKET NUMBER: 1038-681 MIS:jb
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (416) 595-1155
/ TELEFAX: (416) 595-1163
/ INFORMATION FOR SEQ ID NO: 24:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 944 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/
/ US-08-867-941-24
/
/ Query Match 41.6%; Score 42; DB 2; Length 944;
/ Best Local Similarity 60.0%; Pred. No. 1.2e+02;
/ Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
/
/ QY 5 VRLALSGFDGRVGS 19
/ Db 127 VQSLSGYGGGSG 141
/
/ RESULT 22
/ US-09-074-658-23
/ Sequence 23, Application US/09074658
/ Patent No. 6184371
/ GENERAL INFORMATION:
/ APPLICANT: Loosmore, Sheena M
/ APPLICANT: Run-Pan Du
/ APPLICANT: Quijun Wang
/ APPLICANT: Yang, Yan-Ping
/ APPLICANT: Klein, Michel H
/ TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA
/ NUMBER OF SEQUENCES: 78
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Sim & McBurney
```

```
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/074,658
; FILING DATE: 08-MAY-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-795
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 944 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-074-658-23

Query Match 41.6%; Score 42; DB 3; Length 944;
Best Local Similarity 60.0%; Pred. No. 1.2e+02;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 5 VRLALSGFDGRVGS 19
Db 127 VQGLSLGYGGRGSG 141

RESULT 23
US-09-074-658-24
; Sequence 24, Application US/09074658
; Patent No. 6184371
; GENERAL INFORMATION:
; APPLICANT: Loomore, Sheena M
; APPLICANT: Run-Pan Du
; APPLICANT: Quijun Wang
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA
; NUMBER OF SEQUENCES: 78
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/074,658
; FILING DATE: 08-MAY-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-795
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163

; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 944 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-074-658-24

Query Match 41.6%; Score 42; DB 3; Length 944;
Best Local Similarity 60.0%; Pred. No. 1.2e+02;
Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 5 VRLALSGFDGRVGS 19
Db 127 VQGLSLGYGGRGSG 141

RESULT 24
US-08-188-582-14
; Sequence 14, Application US/08188582
; Patent No. 5534410
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comai, Lucio
; APPLICANT: Dynlacht, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Siegfried
; APPLICANT: Tanese, Naoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/188,582
; FILING DATE: 28-JAN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1872 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-188-582-14

Query Match 41.6%; Score 42; DB 1; Length 1872;
Best Local Similarity 56.2%; Pred. No. 2.6e+02;
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 4 AVRLALSGFDGRVGS 19
Db 1740 AQLSESGSDSDVGS 1755
```

RESULT 25

US-08-646-715-14
; Sequence 14, Application US/08646715
; Patent No. 5637686
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comai, Lucio
; APPLICANT: Dynlacht, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Siegfried
; APPLICANT: Tanese, Naoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/646,715
; FILING DATE: 09-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/188,582
; FILING DATE: 28-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 398-3249
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1872 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-646-715-14
Query Match 41.6%; Score 42; DB 1; Length 1872;
Best Local Similarity 56.2%; Pred. No. 2.6e+02;
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 4 AVRLALSGPDGRVGS 19
Db 1740 AIQLSESGSDSDVGS 1755
RESULT 26
US-08-188-582-11
; Sequence 11, Application US/08188582
; Patent No. 5534410
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comai, Lucio
; APPLICANT: Dynlacht, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Siegfried

; APPLICANT: Tanese, Naoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/188,582
; FILING DATE: 28-JAN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1893 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-188-582-11
Query Match 41.6%; Score 42; DB 1; Length 1893;
Best Local Similarity 56.2%; Pred. No. 2.6e+02;
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 4 AVRLALSGPDGRVGS 19
Db 1761 AIQLSESGSDSDVGS 1776
RESULT 27
US-08-646-715-11
; Sequence 11, Application US/08646715
; Patent No. 5637686
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comai, Lucio
; APPLICANT: Dynlacht, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Siegfried
; APPLICANT: Tanese, Naoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent in Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/646,715
/ FILING DATE: 09-MAY-1996
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/188,582
/ FILING DATE: 28-JAN-1994
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Osman, Richard A.
/ REGISTRATION NUMBER: 36,627
/ REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (415) 781-1989
/ TELEFAX: (415) 398-3249
/ TELEX: 910 277299
/ INFORMATION FOR SEQ ID NO: 11:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1893 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-646-715-11

Query Match 41.6%; Score 42; DB 1; Length 1893;
Best Local Similarity 56.2%; Pred. No. 2.6e+02;
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 4 AVRLLALSGFDGRVGS 19
   |::|::|::|::|::|
Db 1761 AIQLSESGSDVDVGS 1776

RESULT 28
US-09-711-164-444
; Sequence 444, Application US/097111164
; Patent No. 6589738
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERAPY
; FILE REFERENCE: ELIPRA.008A
; CURRENT APPLICATION NUMBER: US/09/711,164
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/164415
; PRIOR FILING DATE: 1999-11-9
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 444
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-711-164-444

Query Match 41.1%; Score 41.5; DB 4; Length 471;
Best Local Similarity 55.6%; Pred. No. 62;
Matches 10; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

QY 3 RAVRLALSG-FDGRVGS 19
   |::|::|::|::|::|
Db 261 RTFSLGLAGSFAGRIGS 278

RESULT 29
US-09-252-991A-32206
; Sequence 32206, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

```


Db 41 LSYAVSSGPGPKRGQERC 60

Search completed: May 13, 2004, 06:56:39
Job time : 15 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2004, 06:55:22 ; Search time 214 Seconds
(without alignments)
26.006 Million cell updates/sec

Title: 09549186-8

Perfect score: 101

Sequence: 1 LQRAVRLALSGDFRVSQC 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1145568 seqs, 278261457 residues

Total number of hits satisfying chosen parameters: 1145568

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48	47.5	299	15	US-10-369-493-11875
2	47	46.5	359	9	US-09-815-242-5116
3	47	46.5	359	12	US-10-282-122A-43489
4	46	45.5	162	12	US-10-425-114-50038
5	46	45.5	233	12	US-10-425-114-72819
6	46	45.5	247	12	US-10-425-114-38641
7	46	45.5	404	12	US-10-425-114-71851
8	45.5	45.0	229	9	US-09-738-626-6612
9	45	44.6	263	12	US-10-282-122A-60790
10	45	44.6	490	15	US-10-369-493-15077
11	45	44.6	499	15	US-10-369-493-11671
12	45	44.6	499	15	US-10-369-493-14468
13	44	43.6	184	15	US-10-369-493-3741
14	44	43.6	273	15	US-10-369-493-1170
15	43.5	43.1	300	14	US-10-306-762-80
					Sequence 11875, A
					Sequence 5116, Ap
					Sequence 43489, A
					Sequence 50038, A
					Sequence 72819, A
					Sequence 38641, A
					Sequence 71851, A
					Sequence 6612, Ap
					Sequence 60790, A
					Sequence 15077, A
					Sequence 11671, A
					Sequence 14468, A
					Sequence 3741, Ap
					Sequence 1170, Ap
					Sequence 80, Appl

Sequence 10603, A
Sequence 8867, Ap
Sequence 76545, A
Sequence 12847, A
Sequence 66390, A
Sequence 305, App
Sequence 1228, Ap
Sequence 20594, A
Sequence 17043, A
Sequence 12749, A
Sequence 71581, A
Sequence 1447, Ap
Sequence 642, Ap
Sequence 1302, Ap
Sequence 226329, A
Sequence 1026, Ap
Sequence 534, App
Sequence 277, App
Sequence 14753, A
Sequence 284, App
Sequence 444, App
Sequence 20550, A
Sequence 268979, A
Sequence 190123, A
Sequence 1888, Ap
Sequence 1544, Ap
Sequence 72320, A
Sequence 42523, A
Sequence 4911, Ap
Sequence 10687, A
Sequence 42184, A
Sequence 406, App
Sequence 406, App
Sequence 197, App
Sequence 107, App
Sequence 11394, A
Sequence 5211, Ap
Sequence 138, App
Sequence 71801, A
Sequence 752, App
Sequence 148728, A
Sequence 3482, Ap
Sequence 11907, A
Sequence 66461, A
Sequence 551, App
Sequence 227312, A
Sequence 1083, Ap
Sequence 6, Appli
Sequence 6, Appli
Sequence 6137, Ap
Sequence 447, App
Sequence 17070, A
Sequence 19672, A
Sequence 12248, A
Sequence 4632, Ap
Sequence 50703, A
Sequence 60241, A
Sequence 720, App
Sequence 762, App
Sequence 2526, Ap
Sequence 144979, A
Sequence 59, Appl
Sequence 59, Appl
Sequence 15951, A
Sequence 16315, A
Sequence 19254, A
Sequence 53896, A
Sequence 10173, A
Sequence 9961, Ap
Sequence 3830, Ap
Sequence 3, Appli
Sequence 2100, Ap
Sequence 12957, A
Sequence 54584, A

89 39.5 39.1 335 12 US-10-282-122A-71734 Sequence 71734, A
 90 39.5 39.1 336 9 US-09-815-242-5618 Sequence 5618, Ap
 91 39.5 39.1 336 9 US-09-815-242-12540 Sequence 12540, A
 92 39.5 39.1 336 12 US-10-282-122A-43997 Sequence 43997, A
 93 39.5 39.1 336 15 US-10-369-493-9789 Sequence 9789, Ap
 94 39.5 39.1 336 15 US-10-369-493-19238 Sequence 19238, A
 95 39.5 39.1 337 12 US-10-282-122A-63561 Sequence 63561, A
 96 39.5 39.1 337 12 US-10-282-122A-64191 Sequence 64191, A
 97 39.5 39.1 338 12 US-10-282-122A-73056 Sequence 73056, A
 98 39.5 39.1 418 14 US-09-741-152-6 Sequence 6, Appli
 99 39.5 39.1 418 14 US-10-195-517-6 Sequence 6, Appli
 100 39.5 39.1 762 12 US-10-282-122A-45176 Sequence 45176, A

ALIGNMENTS

RESULT 1
 US-10-369-493-11875 Query Match 47.5%; Score 48; DB 15; Length 299;
 ; Sequence 11875, Application US/10369493 Best Local Similarity 44.4%; Pred. No. 20;
 ; Publication No. US20030233675A1 Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; FILE REFERENCE: 38-10(52053)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28
 ; PRIOR APPLICATION NUMBER: US 60/360,039
 ; PRIOR FILING DATE: 2002-02-21
 ; NUMBER OF SEQ ID NOS: 47374
 ; SEQ ID NO 11875
 ; LENGTH: 299
 ; TYPE: PRT
 ; ORGANISM: Mesorhizobium loti
 US-10-369-493-11875

Query Match 47.5%; Score 48; DB 15; Length 299;
 Best Local Similarity 44.4%; Pred. No. 20;
 Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 LQRAVRLALSGDGRVGS 18
 | | | | | | | | | | | | | | | | | | | | | |
 Db 5 LNRSTRVIVGGFTGKIGS 22

RESULT 2
 US-09-815-242-5116 Query Match 47.5%; Score 48; DB 15; Length 299;
 ; Sequence 5116, Application US/09815242 Best Local Similarity 44.4%; Pred. No. 20;
 ; Patent No. US20020061569A1 Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;
 ; GENERAL INFORMATION:
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari L.
 ; APPLICANT: Zyskind, Judith W.
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John D.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: Identification of Essential Genes in
 ; FILE REFERENCE: ELITRA.011A
 ; CURRENT APPLICATION NUMBER: US/09/815,242
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; NUMBER OF SEQ ID NOS: 14110
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 5116
 ; LENGTH: 359
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-815-242-5116

Query Match 46.5%; Score 47; DB 9; Length 359;
 Best Local Similarity 62.5%; Pred. No. 35;
 Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 LQRAVRLALSGDGRV 16
 | | | | | | | | | | | | | | | | | | | | | |
 Db 126 LDRSARLSLSGIDPRV 141

RESULT 3
 US-10-282-122A-43489 Query Match 46.5%; Score 47; DB 9; Length 359;
 ; Sequence 43489, Application US/10282122A Best Local Similarity 62.5%; Pred. No. 35;
 ; Publication No. US20040029129A1 Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 ; GENERAL INFORMATION:
 ; APPLICANT: Wang, Lianguo
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Carr, Grant
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, R.
 ; APPLICANT: Xu, H.
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 ; FILE REFERENCE: ELITRA.034A
 ; CURRENT APPLICATION NUMBER: US/10/282,122A
 ; CURRENT FILING DATE: 2003-02-20
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/230,335
 ; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: 60/230,347
 ; PRIOR FILING DATE: 2000-09-09
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR APPLICATION NUMBER: 60/267,636
 ; PRIOR FILING DATE: 2001-02-09
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 78614
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 43489
 ; LENGTH: 359
 ; TYPE: PRT

Query Success 43.2%; Score 40; 25 12; 2000 2000
Best Local Similarity 64.3%; Pred. No. 32;

```

RESULT 8
US-09-738-626-6612
; Sequence 6612, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OKAZKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6612
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6612

Query Match          45.0%; Score 45.5; DB 9; Length 229;
Best Local Similarity 47.8%; Pred. No. 37;
Matches 10; Conservative 4; Mismatches 4; Indels 3; Gaps 1;

Qy      2 QRAVRLALSGF---DGRVGSG 19
       ||| : | : | : | : | : | : |
Db     168 QRATQAFTGTLTGDRLCGC 188

RESULT 9
US-10-282-122A-60790
; Sequence 60790, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06

RESULT 10
US-10-369-493-15077
; Sequence 15077, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 15077
; LENGTH: 490
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-10-369-493-15077

Query Match          44.6%; Score 45; DB 15; Length 490;
Best Local Similarity 53.3%; Pred. No. 1e+02;
Matches 8; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy      5 VRLALSGFDGRVGS 19
       :||:|||||:|
Db     440 LHLAMAGYDLRVGN 454

RESULT 11
US-10-369-493-11671
; Sequence 11671, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng

```

```
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 11671
; LENGTH: 499
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-10-369-493-11671

Query Match      44.6%; Score 45; DB 15; Length 499;
Best Local Similarity 53.3%; Pred. No. 1e+02;
Matches 8; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY      5 VRLALSGFDGRVSG 19
DB      449 LHLAMAGYDLRVGNG 463

RESULT 12
US-10-369-493-14468
; Sequence 14468, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 14468
; LENGTH: 499
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-10-369-493-14468

Query Match      44.6%; Score 45; DB 15; Length 499;
Best Local Similarity 53.3%; Pred. No. 1e+02;
Matches 8; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY      5 VRLALSGFDGRVSG 19
DB      449 LHLAMAGYDLRVGNG 463

RESULT 13
US-10-369-493-3741
; Sequence 3741, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 11671
; LENGTH: 499
; TYPE: PRT
; ORGANISM: Agrobacterium tumefaciens
US-10-369-493-3741

Query Match      44.6%; Score 45; DB 15; Length 499;
Best Local Similarity 53.3%; Pred. No. 1e+02;
Matches 8; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY      5 VRLALSGFDGRVSG 19
DB      449 LHLAMAGYDLRVGNG 463

RESULT 14
US-10-369-493-1170
; Sequence 1170, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1170
; LENGTH: 273
; TYPE: PRT
; ORGANISM: Methanobacterium thermoautotrophicum
US-10-369-493-1170

Query Match      43.6%; Score 44; DB 15; Length 273;
Best Local Similarity 53.3%; Pred. No. 78;
Matches 8; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY      5 VRLALSGFDGRVSG 19
DB      2 IRVAVTGACGRMGSG 16

RESULT 15
US-10-306-762-80
; Sequence 80, Application US/10306762
; Publication No. US20030187220A1
; GENERAL INFORMATION:
; APPLICANT: Park, Frances
; APPLICANT: Gajiwala, Ketan S.
; APPLICANT: Buchanan, Sean Grant
; APPLICANT: Sauder, J. Michael
; TITLE OF INVENTION: CRYSTALS AND STRUCTURES OF A FLAVIN
; FILE OF INVENTION: MONONUCLEOTIDE BINDING PROTEIN (FMNBP)
; FILE REFERENCE: 52498-20011.00
; CURRENT APPLICATION NUMBER: US/10/306,762
; CURRENT FILING DATE: 2003-04-16
; PRIOR FILING DATE: 2003-04-16
; PRIOR APPLICATION NUMBER: US 60/334,132
; PRIOR FILING DATE: 2001-11-28
; NUMBER OF SEQ ID NOS: 242
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 80
; LENGTH: 300
```

```
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 3741
; LENGTH: 184
; TYPE: PRT
; ORGANISM: Neurospora crassa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(184)
; OTHER INFORMATION: unsure at all xaa locations
US-10-369-493-3741

Query Match      43.6%; Score 44; DB 15; Length 184;
Best Local Similarity 47.4%; Pred. No. 51;
Matches 9; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY      2 QRAVRLALSGFDGRVSGC 20
DB      163 EAATRAALLTFEKEGSGC 181

RESULT 14
US-10-369-493-1170
; Sequence 1170, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1170
; LENGTH: 273
; TYPE: PRT
; ORGANISM: Methanobacterium thermoautotrophicum
US-10-369-493-1170

Query Match      43.6%; Score 44; DB 15; Length 273;
Best Local Similarity 53.3%; Pred. No. 78;
Matches 8; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY      5 VRLALSGFDGRVSG 19
DB      2 IRVAVTGACGRMGSG 16

RESULT 15
US-10-306-762-80
; Sequence 80, Application US/10306762
; Publication No. US20030187220A1
; GENERAL INFORMATION:
; APPLICANT: Park, Frances
; APPLICANT: Gajiwala, Ketan S.
; APPLICANT: Buchanan, Sean Grant
; APPLICANT: Sauder, J. Michael
; TITLE OF INVENTION: CRYSTALS AND STRUCTURES OF A FLAVIN
; FILE OF INVENTION: MONONUCLEOTIDE BINDING PROTEIN (FMNBP)
; FILE REFERENCE: 52498-20011.00
; CURRENT APPLICATION NUMBER: US/10/306,762
; CURRENT FILING DATE: 2003-04-16
; PRIOR FILING DATE: 2003-04-16
; PRIOR APPLICATION NUMBER: US 60/334,132
; PRIOR FILING DATE: 2001-11-28
; NUMBER OF SEQ ID NOS: 242
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 80
; LENGTH: 300
```

; TYPE: PRT
; ORGANISM: G. diazotrophicus (20257536)
US-10-306-762-80

Query Match 43.1%; Score 43.5; DB 14; Length 300;
Best Local Similarity 66.7%; Pred. No. 1e+02;
Matches 10; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 7 LALSGFDG-RVGGC 20
DB 194 LALSGADGVMIGRC 208

RESULT 16

US-10-156-761-10603
; Sequence 10603, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10603
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10603

Query Match 42.6%; Score 43; DB 14; Length 147;
Best Local Similarity 64.3%; Pred. No. 58;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 QRAVRLALSGFDR 15
DB 39 ERAERLSQYFDR 52

RESULT 17

US-10-369-493-8867
; Sequence 8867, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 8867
; LENGTH: 261
; TYPE: PRT
; ORGANISM: Chloroflexus aurantiacus
US-10-369-493-8867

Query Match 42.6%; Score 43; DB 15; Length 261;
Best Local Similarity 47.4%; Pred. No. 1.1e+02;
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 LQRAVRLALSGFDRVGG 19
DB 92 MRGVNALGLGDLGRLLSG 110

RESULT 18

US-10-282-122A-76545
; Sequence 76545, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 76545
; LENGTH: 287
; TYPE: PRT
; ORGANISM: Treponema pallidum
US-10-282-122A-76545

Query Match 42.6%; Score 43; DB 12; Length 287;
Best Local Similarity 58.3%; Pred. No. 1.2e+02;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 9 LSGFDRVGGC 20
DB 193 ICGFAGSVGAGC 204

RESULT 19

US-10-156-761-12847
; Sequence 12847, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:


```
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12847
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-12847

Query Match 42.6%; Score 43; DB 14; Length 383;
Best Local Similarity 43.8%; Pred. No. 1.6e+02;
Matches 7; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 2 QRAVRLALSGFDGRVG 17
::|||::|||
Db 3 RKTVAIANGVTGRMG 18

RESULT 20
US-10-282-122A-66390
; Sequence 66390, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 66390
; LENGTH: 1221
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-66390

Query Match 42.6%; Score 43; DB 12; Length 1221;
Best Local Similarity 52.9%; Pred. No. 5.5e+02;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 LQRAVRLALSGFDGRVG 17
|||::|||::|||
Db 36 LGRVFLQVSGFGRIG 52

RESULT 21
US-09-764-872-305
; Sequence 305, Application US/09764872
; Publication No. US20030050231A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA125
; CURRENT APPLICATION NUMBER: US/09/764,872
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 957
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 305
; LENGTH: 78
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-872-305

Query Match 42.1%; Score 42.5; DB 10; Length 78;
Best Local Similarity 52.2%; Pred. No. 35;
Matches 12; Conservative 1; Mismatches 7; Indels 3; Gaps 1;

QY 1 LQRAVRLAL--SGFDGRVSGC 20
|||::|||::|||
Db 36 LQSVSLALCHSPSSVDGWLGGSC 58

RESULT 22
US-10-369-493-1228
; Sequence 1228, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1228
; LENGTH: 301
; TYPE: PRT
; ORGANISM: Methanobacterium thermoautotrophicum
US-10-369-493-1228

Query Match 42.1%; Score 42.5; DB 15; Length 301;
Best Local Similarity 50.0%; Pred. No. 1.5e+02;
Matches 10; Conservative 2; Mismatches 5; Indels 3; Gaps 1;
```

QY 3 RAVRLALSGDGR---VGSG 19
Db 133 QTIREKLGSGDGRLLVFGDG 152

RESULT 23

US-10-369-493-20594
; Sequence 20594, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 20594
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Rhodospseudomonas palustris
US-10-369-493-20594

Query Match 42.1%; Score 42.5; DB 15; Length 331;
Best Local Similarity 64.3%; Pred. No. 1.7e+02;
Matches 9; Conservative 4; Mismatches 0; Indels 1; Gaps 1;

QY 4 AVRRLALSGDGRVG 17
Db 2 AVRVAINGF-GRIG 14

RESULT 24

US-10-369-493-17043
; Sequence 17043, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 17043
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Caulobacter crescentus
US-10-369-493-17043

Query Match 42.1%; Score 42.5; DB 15; Length 335;
Best Local Similarity 64.3%; Pred. No. 1.7e+02;
Matches 9; Conservative 4; Mismatches 0; Indels 1; Gaps 1;

QY 4 AVRRLALSGDGRVG 17
Db 2 AVRVAINGF-GRIG 14

RESULT 25

US-10-156-761-12749

; Sequence 12749, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12749
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-12749

Query Match 41.6%; Score 42; DB 14; Length 215;
Best Local Similarity 57.1%; Pred. No. 1.3e+02;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 5 VRLALSGDGRVGS 18
Db 43 IRLAARGCDGRGT 56

RESULT 26

US-10-425-114-71581
; Sequence 71581, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 71581
; LENGTH: 222
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLMO17160B05_FLI.pep
US-10-425-114-71581

Query Match 41.6%; Score 42; DB 12; Length 222;
Best Local Similarity 56.2%; Pred. No. 1.3e+02;
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 4 AVRRLALSGDGRVGS 19
Db 75 ALSLASNGEGRVPVG 90

RESULT 27

US-10-412-699B-1447
; Sequence 1447, Application US/10412699B
; Publication No. US20040045049A1
; GENERAL INFORMATION:

APPLICANT: Mendel Biotechnology, Inc.
APPLICANT: Zhang, James
APPLICANT: Fromm, Michael E.
APPLICANT: Heard, Jacqueline E.
APPLICANT: Riechmann, Jose Luis
APPLICANT: Adam, Luc J.
APPLICANT: Broun, Pierre E.
APPLICANT: Pineda, Omaira
APPLICANT: Reuber, T. Lynne
APPLICANT: Keddle, James S.
APPLICANT: Yu, Guo-Liang
APPLICANT: Jiang, Cai-Zhong
APPLICANT: Samaha, Raymond R.
APPLICANT: Pilgrim, Marsha L.
APPLICANT: Creelman, Robert A.
APPLICANT: DuBell, Arnold N.
APPLICANT: Ratcliffe, Oliver
APPLICANT: Kumimoto, Roderick
APPLICANT: Sherman, Bradley K.
TITLE OF INVENTION: Polynucleotides and Polypeptides in Plants
FILE REFERENCE: MBI-0048CIP
CURRENT APPLICATION NUMBER: US/10/412,699B
CURRENT FILING DATE: 2003-04-10
PRIOR FILING DATE: 1999-09-13
PRIOR APPLICATION NUMBER: 09/394,519
PRIOR FILING DATE: 2000-01-21
PRIOR APPLICATION NUMBER: 09/489,376
PRIOR FILING DATE: 2000-02-17
PRIOR APPLICATION NUMBER: 09/506,720
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 09/533,030
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 09/533,392
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 09/533,029
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 09/532,591
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 09/533,648
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 09/713,994
PRIOR FILING DATE: 2000-11-16
PRIOR APPLICATION NUMBER: 09/819,142
PRIOR FILING DATE: 2001-03-27
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 2011
SOFTWARE: PatentIn version 3.2
SEQ ID NO 1447
LENGTH: 243
TYPE: PRT
ORGANISM: Oryza sativa
US-10-412-699B-1447

Query Match 41.6%; Score 42; DB 12; Length 243;
Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 8 ALSGFDGRVGG 19
DB 78 AVAGRDGRVGG 89

RESULT 28

US-10-310-154-642
Sequence 642, Application US/10310154
Publication No. US20030233670A1
GENERAL INFORMATION:
APPLICANT: Edgerton, Michael D
APPLICANT: Chomet, Paul S.
APPLICANT: Adams, Thomas H
APPLICANT: Ruff, Thomas G.
APPLICANT: Agarwal, Ameeta K.
APPLICANT: Ahrens, Jeffrey E.
APPLICANT: Ball, James A.

APPLICANT: Banu, G.
APPLICANT: Bell, Erin
APPLICANT: Boddupalli, Raghava
APPLICANT: Deikman, Jill
APPLICANT: Deng, Molian
APPLICANT: Dong, Jinzhao M.
APPLICANT: Duff, Stephen M.
APPLICANT: Galligan, Meghan M.
APPLICANT: Hinchey, Brenda S.
APPLICANT: Huang, Shihshieh
APPLICANT: Johnson, G. Richard
APPLICANT: Jung, Vincent
APPLICANT: Kretzmer, Keith A
APPLICANT: Laccetti, Lucille B.
APPLICANT: Lai, Chao-Qiang
APPLICANT: Lee, Gary
APPLICANT: Lin, Jie-Yi
APPLICANT: Liu, Jingdong
APPLICANT: Lu, Bin
APPLICANT: Luethy, Michael M.
APPLICANT: Lund, Adrian
APPLICANT: Madson, Linda L.
APPLICANT: Malloy, Kathleen A.
APPLICANT: McKiel, Christine L.
APPLICANT: Miller, Philip W.
APPLICANT: Padmavathi, Manchikanti
APPLICANT: Parnell, Laurence D.
APPLICANT: Start, William G.
APPLICANT: Tennesen, Dan
APPLICANT: Vidya, K.R.
APPLICANT: Wang, Haiyun
APPLICANT: Xin, Zhanqun
APPLICANT: Xu, Nanfei
APPLICANT: Yang, Chunzhi
APPLICANT: Zeng, Xiaoping
APPLICANT: Zhang, Qiang
APPLICANT: Zhao, Yajuan
APPLICANT: Zhou, Li
TITLE OF INVENTION: Gene Sequences and Uses Thereof in Plants
FILE REFERENCE: 38-15(52796)B
CURRENT APPLICATION NUMBER: US/10/310,154
CURRENT FILING DATE: 2002-12-04
PRIOR APPLICATION NUMBER: 60/337,358
PRIOR FILING DATE: 2001-12-04
NUMBER OF SEQ ID NOS: 736
SEQ ID NO 642
LENGTH: 243
TYPE: PRT
ORGANISM: Oryza sativa
US-10-310-154-642

Query Match 41.6%; Score 42; DB 15; Length 243;
Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 8 ALSGFDGRVGG 19
DB 78 AVAGRDGRVGG 89

RESULT 29

US-10-374-780A-1302
Sequence 1302, Application US/10374780A
Publication No. US20040019927A1
GENERAL INFORMATION:
APPLICANT: Sherman, Bradley K
APPLICANT: Riechmann, Jose Luis
APPLICANT: Jiang, Cai-Zhong
APPLICANT: Heard, Jacqueline E
APPLICANT: Haake, Volker
APPLICANT: Creelman, Robert A
APPLICANT: Ratcliffe, Oliver
APPLICANT: Adam, Luc J

APPLICANT: Reuber, T. Lynne
 APPLICANT: Keddle, James
 APPLICANT: Brown, Pierre E
 APPLICANT: Filgrim, Marsha L
 APPLICANT: Dubell III, Arnold T
 APPLICANT: Pineda, Omaira
 APPLICANT: Yu, Guo-Liang
 TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
 FILE REFERENCE: MBI-0047 CIP
 CURRENT APPLICATION NUMBER: US/10/374,780A
 CURRENT FILING DATE: 2003-02-25
 PRIOR APPLICATION NUMBER: 09/837,944
 PRIOR FILING DATE: 2001-04-18
 PRIOR APPLICATION NUMBER: 60/310,847
 PRIOR FILING DATE: 2001-08-09
 PRIOR APPLICATION NUMBER: 09/934,455
 PRIOR FILING DATE: 2001-08-22
 PRIOR APPLICATION NUMBER: 60/336,049
 PRIOR FILING DATE: 2001-11-19
 PRIOR APPLICATION NUMBER: 60/338,692
 PRIOR FILING DATE: 2001-12-11
 PRIOR APPLICATION NUMBER: 10/171,468
 PRIOR FILING DATE: 2002-06-14
 PRIOR APPLICATION NUMBER: 10/225,066
 PRIOR FILING DATE: 2002-08-09
 PRIOR APPLICATION NUMBER: 10/225,067
 PRIOR FILING DATE: 2002-08-09
 PRIOR APPLICATION NUMBER: 10/225,068
 PRIOR FILING DATE: 2002-08-09
 NUMBER OF SEQ ID NOS: 2906
 SOFTWARE: PatentIn version 3.2
 SEQ ID NO 1302
 LENGTH: 243
 TYPE: PRT
 ORGANISM: Oryza sativa
 FEATURE:
 OTHER INFORMATION: Orthologous to G975, G2583
 US-10-374-780A-1302

Query Match 41.6%; Score 42; DB 15; Length 243;
 Best Local Similarity 66.7%; Pred. No. 1.4e+02;
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 8 ALSGPDGRVGS 19
 DB 78 AVAGRDGRVGG 89

RESULT 30
 US-10-424-599-226329
 Sequence 226329, Application US/10424599
 Publication No. US20040031072A1
 GENERAL INFORMATION:
 APPLICANT: La Rosa Thomas J
 APPLICANT: Kovalic David K
 APPLICANT: Zhou Yihua
 APPLICANT: Cao Yongwei
 TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
 TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 FILE REFERENCE: 38-21(53223)B
 CURRENT APPLICATION NUMBER: US/10/424,599
 CURRENT FILING DATE: 2003-04-28
 NUMBER OF SEQ ID NOS: 285684
 SEQ ID NO 226329
 LENGTH: 438
 TYPE: PRT
 ORGANISM: Glycine max
 FEATURE:
 NAME/KEY: unsure
 LOCATION: (1)..(438)
 OTHER INFORMATION: unsure at all Xaa locations
 FEATURE:
 OTHER INFORMATION: Clone ID: PAT_MRT3847_46404C.1.pap

US-10-424-599-226329
 Query Match 41.6%; Score 42; DB 12; Length 438;
 Best Local Similarity 43.8%; Pred. No. 2.7e+02;
 Matches 7; Conservative 6; Mismatches 3; Indels 0; Gaps 0;
 QY 3 RAVRLALSGFDGRVGS 18
 DB 114 RXMKLVVTGYEGFVGS 129
 Search completed: May 13, 2004, 07:17:03
 Job time : 215 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2004, 06:49:47 ; Search time 12 Seconds
(without alignments)
160.319 Million cell updates/sec

Title: 09549186-8

Perfect score: 101

Sequence: 1 LQRAVRLALSGDGRVSGSC 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : PIR 78:**

1: pir1:**
2: pir2:**
3: pir3:**
4: pir4:**

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47	46.5	359	E83312	hypothetical prote
2	47	46.5	575	D69512	acetolactate synth
3	45.5	45.0	473	D83611	probable transcrip
4	45	44.6	263	AC1313	dihydrodipicolinat
5	45	44.6	367	A95322	NolF secretion pro
6	45	44.6	503	AB2938	alpha-L-arabinofur
7	45	44.6	503	E98344	hypothetical prote
8	44	43.6	136	E95333	hypothetical prote
9	44	43.6	273	H62006	dihydrodipicolinat
10	44	43.6	367	S16562	nolF protein - Rhi
11	43.5	43.1	585	E70533	probable glpD2 pro
12	43.5	43.1	693	H64139	DNA helicase recG
13	43	42.6	287	D71304	probable phosphati
14	43	42.6	352	AB1682	chitinases homolog
15	43	42.6	352	AC1310	chitinases homolog
16	43	42.6	377	H83120	conserved hypothet
17	43	42.6	477	H84222	hypothetical prote
18	43	42.6	1221	E83327	conserved hypothet
19	42.5	42.1	301	E69059	ornithine carbamoy
20	42.5	42.1	335	F87651	glyceraldehyde 3-p
21	42.5	42.1	338	AC0290	glyceraldehyde-3-p
22	42.5	42.1	338	AC0113	D-erythrose 4-phos
23	42	41.6	135	E98300	conserved hypothet
24	42	41.6	135	E98300	hypothetical prote
25	42	41.6	263	AC1685	dihydrodipicolinat
26	42	41.6	450	DCCHO	ornithine decarbox
27	42	41.6	466	T45585	hypothetical prote
28	42	41.6	488	A95926	probable argininos
29	42	41.6	862	T34342	hypothetical prote

30	42	41.6	940	2	S49087	lactoferrin bindin
31	42	41.6	943	2	G81070	lactoferrin-bindin
32	42	41.6	944	2	C81798	reverse transcript
33	42	41.6	1151	2	T30936	hypothetical prote
34	42	41.6	1545	2	F98262	glutamate synthase
35	42	41.6	1581	2	AD3022	transcription init
36	42	41.6	1865	1	I48155	transcription init
37	42	41.6	1893	1	A40262	hypothetical prote
38	41.5	41.1	471	2	D64974	probable transport
39	41.5	41.1	471	2	E90989	probable transport
40	41.5	41.1	471	2	H85834	comM protein - The
41	41.5	41.1	501	2	G72368	hypothetical prote
42	41.5	41.1	689	2	T08918	hypothetical prote
43	41	40.6	100	2	T48765	ferredoxin [limpo
44	41	40.6	106	2	AF2742	ferredoxin, 2Fe-2S
45	41	40.6	143	2	E97523	probable dehydroge
46	41	40.6	251	2	S25186	heterogeneous ribo
47	41	40.6	282	2	A31765	probable epiA prot
48	41	40.6	322	1	C70714	hypothetical prote
49	41	40.6	425	2	T14622	carbon starvation
50	41	40.6	592	2	E75032	tegument protein 0
51	41	40.6	1319	2	S55598	glyceraldehyde 3-p
52	40.5	40.1	334	2	H83246	glyceraldehyde-3-p
53	40.5	40.1	374	2	T07990	carbamoyl-phosphat
54	40.5	40.1	1079	2	F84946	hypothetical prote
55	40	39.6	142	2	E87334	chromosome partiti
56	40	39.6	203	2	C75366	hypothetical prote
57	40	39.6	226	2	D71165	hypothetical prote
58	40	39.6	229	1	BVECHJ	molybdenum transpo
59	40	39.6	229	2	H90727	molybdate transpor
60	40	39.6	229	2	A85579	molybdenum transpo
61	40	39.6	229	2	AD0595	dihydrodipicolinat
62	40	39.6	257	2	D87689	probable integral
63	40	39.6	264	2	T35056	probable oxidoredu
64	40	39.6	279	2	T34860	prephenate dehydro
65	40	39.6	281	2	AB1949	transcription regu
66	40	39.6	309	2	AC1182	genome polyprotein
67	40	39.6	330	2	S47491	probable molybdopt
68	40	39.6	355	2	G72743	conserved hypothet
69	40	39.6	402	2	G83289	molybdenum cofacto
70	40	39.6	407	2	D84169	ornithine decarbox
71	40	39.6	431	2	S64704	ornithine decarbox
72	40	39.6	433	2	T03632	probable DNA-bindi
73	40	39.6	433	2	T03035	hypothetical prote
74	40	39.6	439	2	T52291	hypothetical prote
75	40	39.6	445	2	H6560	ABC transport prot
76	40	39.6	455	2	T05081	probable ABC trans
77	40	39.6	1218	2	AD0837	glyceraldehyde-3-p
78	40	39.6	1218	2	T30293	glyceraldehyde-3-p
79	39.5	39.1	331	2	S57280	glyceraldehyde-3-p
80	39.5	39.1	331	2	S57281	glyceraldehyde-3-p
81	39.5	39.1	332	2	C81285	glyceraldehyde-3-p
82	39.5	39.1	336	2	E85850	glyceraldehyde-3-p
83	39.5	39.1	337	1	DE2YG3	glyceraldehyde-3-p
84	39.5	39.1	337	2	S73737	glyceraldehyde-3-p
85	39.5	39.1	348	2	AE0875	D-erythrose 4-phos
86	39.5	39.1	404	2	AF0034	probable integral
87	39.5	39.1	521	2	S54266	glycoprotein GC -
88	39.5	39.1	521	2	S54266	hypothetical prote
89	39.5	39.1	986	2	G65116	cobalamin biosynth
90	39.5	39.1	1248	2	G83278	cobN protein [impo
91	39.5	39.1	1263	2	AI3338	ferredoxin, 2Fe-2S
92	39	38.6	113	2	AI3371	hypothetical prote
93	39	38.6	142	2	AI1975	lmbK protein - Str
94	39	38.6	190	1	S44958	hypothetical prote
95	39	38.6	203	2	G83227	conserved hypothet
96	39	38.6	222	2	AH2619	probable endozepti
97	39	38.6	222	2	H97401	thrombin (PC 3.4.2
98	39	38.6	235	2	D42696	ORF1 protein - sai
99	39	38.6	256	2	A34770	ribosomal protein
100	39	38.6	275	2	T11810	

ALIGNMENTS

RESULT 1	
E83312	
hypothetical protein PA2671 [imported] - Pseudomonas aeruginosa (strain PAO1)	
C:Species: Pseudomonas aeruginosa	
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000	
C:Accession: E83312	
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; Br	
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,	
; Lory, S.; Olson, M.V.	
Nature 406, 959-964, 2000	
A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic patho	
A:Reference number: A82950; MUID:20437337; PMID:10984043	
A:Accession: E83312	
A:Status: preliminary	
A:Molecule type: DNA	
A:Residues: 1-359 <STO>	
A:Cross-references: GB:AE004695; GB:AE004091; NID:g9948730; PIDN:AAG06059.1; GSPDB:GN001	
A:Experimental source: strain PAO1	
C:Genetics:	
A:Gene: PA2671	
Query Match 46.5%; Score 47; DB 2; Length 359;	
Best Local Similarity 62.5%; Pred. No. 9.1;	
Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;	
QY 1 LQRAVRLALSGFGDGRV 16	
Db 126 LDRSARLSLGGIDPRV 141	
RESULT 2	
D69512	
acetylactate synthase, large subunit (ilvB-4) homolog - Archaeoglobus fulgidus	
C:Species: Archaeoglobus fulgidus	
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999	
C:Accession: D69512	
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson	
; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.	
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.	
Nature 390, 364-370, 1997	
A:Authors: Uterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.	
Smith, H.O.; Woese, C.R.; Venter, J.C.	
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo	
A:Reference number: A69250; MUID:98049343; PMID:9389475	
A:Accession: D69512	
A:Status: preliminary; nucleic acid sequence not shown; translation not shown	
A:Molecule type: DNA	
A:Residues: 1-575 <KLE>	
A:Cross-references: GB:AE000958; GB:AE000782; NID:g2689281; PIDN:AAB89146.1; PID:g264842	
C:Superfamily: acetolactate synthase large chain; thiamin pyrophosphate-binding domain h	
Query Match 46.5%; Score 47; DB 2; Length 575;	
Best Local Similarity 62.5%; Pred. No. 15;	
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;	
QY 1 LQRAVRLALSGFGDGRV 16	
Db 147 VQRAIRIALSGRGPV 162	
RESULT 3	
D83611	
probable transcription regulator PA0268 [imported] - Pseudomonas aeruginosa (strain PAO1)	
C:Species: Pseudomonas aeruginosa	
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000	
C:Accession: D83611	
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; Br	
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,	
; Lory, S.; Olson, M.V.	
Nature 406, 959-964, 2000	

A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic patho	
A:Reference number: A82950; MUID:20437337; PMID:10984043	
A:Accession: D83611	
A:Status: preliminary	
A:Molecule type: DNA	
A:Residues: 1-473 <STO>	
A:Cross-references: GB:AE004465; GB:AE004091; NID:g9946107; PIDN:AAG03657.1; GSPDB:GN001;	
A:Experimental source: strain PAO1	
C:Genetics:	
A:Gene: PA0268	
C:Superfamily: hypothetical protein b1439	
Query Match 45.0%; Score 45.5; DB 2; Length 473;	
Best Local Similarity 73.3%; Pred. No. 21;	
Matches 11; Conservative 1; Mismatches 2; Indels 1; Gaps 1;	
QY 6 RLALSGF-DGRVGS 19	
Db 59 RLASEGFLEGRVGS 73	
RESULT 4	
AC1313	
dihydrodipicolinate reductase homolog dapB [imported] - Listeria monocytogenes (strain EX	
C:Species: Listeria monocytogenes	
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001	
C:Accession: AC1313	
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker,	
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duesurget, O.; Entian, K.D.; Fsihi, H.;	
D.; Jones, L.M.; Karst, U.	
Science 294, 849-852, 2001	
A:Authors: Kreft, J.; Kuhn, M.; Kunat, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mat	
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,	
A:Title: Comparative Genomics of Listeria species.	
A:Reference number: AB1077; MUID:21537279; PMID:11679669	
A:Accession: AC1313	
A:Status: preliminary	
A:Molecule type: DNA	
A:Residues: 1-263 <GLA>	
A:Cross-references: GB:NC_003210; PIDN:CAC99985.1; PID:g16411360; GSPDB:GN00177	
A:Experimental source: strain EGB-e	
C:Genetics:	
A:Gene: dapB	
C:Superfamily: dihydrodipicolinate reductase	
Query Match 44.6%; Score 45; DB 2; Length 263;	
Best Local Similarity 61.5%; Pred. No. 14;	
Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;	
QY 5 VRLALSGPDGRVG 17	
Db 1 MRVAVSGFKRGMG 13	
RESULT 5	
A95322	
NoLP secretion protein [imported] - Sinorhizobium meliloti (strain 1021) magaplaamid pSyn	
C:Species: Sinorhizobium meliloti	
C>Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001	
C:Accession: A95322	
R:Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bowse	
; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Welle, D.H.; Yeh, K.C.;	
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001	
A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti	
A:Reference number: A95262; MUID:21396509; PMID:11481432	
A:Accession: A95322	
A:Status: preliminary	
A:Molecule type: DNA	
A:Residues: 1-367 <KUR>	
A:Cross-references: GB:AE006469; PIDN:AAK65139.1; PID:g14523579; GSPDB:GN00165	
A:Experimental source: strain 1021, megaplaamid pSymA	
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, P.; Barloy-Hubler,	
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federpspiel, N.A.; Fisher, R.F.;	

L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure, H.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
 A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
 A:Reference number: A96039; MUID:21368234; PMID:11474104
 A:Contents: annotation
 C:Genetics:
 A:Gene: nolF
 A:Genome: plasmid

Query Match 44.6%; Score 45; DB 2; Length 367;
 Best Local Similarity 47.4%; Pred. No. 19;
 Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 LQRAVRLALSGFGDGRVSG 19
 : |||: |||: |||:
 Db 230 LKQSVLTVGFGPGRFSG 248

RESULT 6

AB2938
 alpha-L-arabinofuranosidase Atu3104 [imported] - Agrobacterium tumefaciens (strain C58,
 C:Species: Agrobacterium tumefaciens
 C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kuttyavin, T.; Levy, R.; Li, M.; McClell
 ; Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 sser, E.W.
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A:Reference number: AB2577; MUID:21608550; PMID:11743193
 A:Accession: AB2938
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-503 <KUR>
 A:Cross-references: GB:AE008689; PIDN:AA43920.1; PID:g17741471; GSPDB:GN00187
 A:Experimental source: strain C58 (Dupont)
 C:Genetics:
 A:Gene: Atu3104
 A:Map position: linear chromosome

Query Match 44.6%; Score 45; DB 2; Length 503;
 Best Local Similarity 53.3%; Pred. No. 27;
 Matches 8; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 5 VRLALSGDGRVSG 19
 : |||: |||: |||:
 Db 450 LHLAMAGYDLRVGNG 464

RESULT 7

E98344
 hypothetical protein AGR_L_3408 [imported] - Agrobacterium tumefaciens (strain C58, Cere
 C:Species: Agrobacterium tumefaciens
 C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
 C:Accession: E98344
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
 Science 294, 2323-2328, 2001
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
 A:Reference number: A97359; MUID:21608551; PMID:11743194
 A:Accession: E98344
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-503 <KUR>
 A:Cross-references: GB:AE007870; PIDN:AAK90279.1; PID:g15160302; GSPDB:GN00170
 C:Genetics:
 A:Gene: AGR_L_3408
 A:Map position: linear chromosome

Query Match 44.6%; Score 45; DB 2; Length 503;
 Best Local Similarity 53.3%; Pred. No. 27;
 Matches 8; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 5 VRLALSGDGRVSG 19
 : |||: |||: |||:
 Db 450 LHLAMAGYDLRVGNG 464

RESULT 8

B95333
 hypothetical protein Sma1053 [imported] - Sinorhizobium meliloti (strain 1021) magaplasmi
 C:Species: Sinorhizobium meliloti
 C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
 C:Accession: B95333
 R:Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bowe
 ; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.;
 Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
 A:Title: Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti*
 A:Reference number: A95262; MUID:21396509; PMID:11481432
 A:Accession: B95333
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-136 <KUR>
 A:Cross-references: GB:AE006469; PIDN:AAK65228.1; PID:g14523677; GSPDB:GN00165
 A:Experimental source: strain 1021, megaplasmid pSymA
 R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
 pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
 L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
 hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.C.
 A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
 A:Reference number: A96039; MUID:21368234; PMID:11474104
 A:Contents: annotation
 C:Genetics:
 A:Gene: Sma1053
 A:Genome: plasmid

Query Match 43.6%; Score 44; DB 2; Length 136;
 Best Local Similarity 66.7%; Pred. No. 10;
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 9 LSGFDGRVSGC 20
 : |||: |||: |||:
 Db 86 LVGFGSRVSGC 97

RESULT 9

H69206
 dihydrodipicolinate reductase - Methanobacterium thermoautotrophicum (strain Delta H)
 C:Species: Methanobacterium thermoautotrophicum
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 11-Jun-1999
 C:Accession: H69206
 R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; E
 ; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiawani, N.;
 ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
 J. Bacteriol. 179, 7135-7155, 1997
 A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: functi
 A:Reference number: A69000; MUID:98037514; PMID:9371463
 A:Accession: H69206
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-273 <MTH>
 A:Cross-references: GB:AE000858; GB:AE000666; NID:g2621885; PIDN:AA885300.1; PID:g2621899

Query Match 43.6%; Score 44; DB 2; Length 273;
 Best Local Similarity 53.3%; Pred. No. 21;

QY 9 LSGFDGRVSGC 20
 : |||: |||: |||:
 Db 86 LVGFGSRVSGC 97

Matches 8; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 5 VRLALSGPGRVSG 19
:|:|:| | | | | |
Db 2 IRVAVTGACGRMGSG 16

RESULT 10
S16562
nolP protein - Rhizobium meliloti
C:Species: Rhizobium meliloti
C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 08-Oct-1999
C:Accession: S16562
R:Baev, N.; Endre, G.; Petrovics, G.; Banfalvi, Z.; Kondorosi, A.
Mol. Gen. Genet. 228, 113-124, 1991
A>Title: Six nodulation genes of nod box locus 4 in Rhizobium meliloti are involved in n
A:Reference number: S16561; MUID:91360053; PMID:1909418
A:Accession: S16562
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-367 <BAE>
A:CROSS-references: EMBL:X58632; NID:946331; PIDN:CAA41486.1; PID:946333

Query Match 43.6%; Score 44; DB 2; Length 367;
Best Local Similarity 47.4%; Pred. No. 28;
Matches 9; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 LQRAVRLALSGPGRVSG 19
:|:|:| | | | | |
Db 229 LKQSVELMVEGPPGRTFSG 247

RESULT 11
H70533
probable glpD2 protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: H70533
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Sgares, R.; Sultston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: H70533
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-585 <COL>
A:CROSS-references: GB:Z96070; GB:AL123456; NID:93261791; PIDN:CAB09447.1; PID:93261792
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: glpD2
C:Superfamily: glycerol-3-phosphate dehydrogenase (aerobic)

Query Match 43.1%; Score 43.5; DB 2; Length 585;
Best Local Similarity 50.0%; Pred. No. 54;
Matches 14; Conservative 0; Mismatches 5; Indels 9; Gaps 2;

QY 2 QRAV---RLALSGFD-----GRVSGGC 20
| | | | | | | | | | | | | |
Db 23 QRAVAKRLGTQFDVVVGGVVGSGC 50

RESULT 12
E64139
DNA helicase recG protein - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001
C:Accession: E64139; T09431
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.

Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, J.
A>Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:93350630; PMID:7542800
A:Accession: E64139
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-693 <TIGR>
A:CROSS-references: GB:U12847; GB:L42023; NID:93212238; PIDN:AAC23387.1; PID:91574599; T0
R.White, O.; Clayton, R.A.; Kerlavage, A.R.; Fleischmann, R.D.; Peterson, J.; Hickey, E.;
submitted to the EMBL Data Library, May 1998
A:Reference number: Z16667
A:Accession: T09431
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-693 <WHI>
A:CROSS-references: EMBL:U32847; NID:93212238; PID:91574599
C:Genetics:
A:Gene: recG; H11740
C:Superfamily: DNA helicase recG
C:Keywords: ATP; DNA binding; nucleotide binding; P-loop
F:296-303/Region: nucleotide-binding motif A (P-loop)
F:393-398/Region: nucleotide-binding motif B
F:397-400/Region: DEXH motif

Query Match 43.1%; Score 43.5; DB 1; Length 693;
Best Local Similarity 47.8%; Pred. No. 65;
Matches 11; Conservative 2; Mismatches 7; Indels 3; Gaps 1;

QY 1 LQRAVRLALSGFD---GRVSGGC 20
:|:|:| | | | | |
Db 567 IENAERLGLSQLHQLRGRVGRGC 589

RESULT 13
D71304
probable phosphatidate cytidylyltransferase (cdsa) - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C>Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 17-Mar-2000
C:Accession: D71304
R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDo
rson, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A>Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A:Reference number: A71250; MUID:98332770; PMID:9665876
A:Accession: D71304
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-287 <COL>
A:CROSS-references: GB:AE001235; GB:AE000520; NID:93322893; PIDN:AAC65575.1; PID:93322896
A:Experimental source: strain Nichols
C:Genetics:
A:Gene: TP0602
C:Superfamily: phosphatidate cytidylyltransferase

Query Match 42.6%; Score 43; DB 2; Length 287;
Best Local Similarity 58.3%; Pred. No. 32;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 9 LSGFDGRVSGGC 20
:|:|:| | | | | |
Db 193 IGGFAGSVGAGC 204

RESULT 14
AB1682
chitinases homolog lin1996 [imported] - Listeria innocua (strain Clip11262)
C:Species: Listeria innocua
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AB1682
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker, A.
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duesurget, O.; Entian, K.D.; Feihl, H.;

D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Maier,
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vasquez-Boland, J.A.; Voss, H.; Wehlend,
A;Title: Comparative genomics of *Listeria* species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AB1882
A;Status: preliminary
A:Molecule type: DNA
A;Residues: 1-352 <GLA>
A;Cross-references: GB:AL592023; PIDN:CAC97226.1; PID:g16414497; GSPDB:GN00178
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: lnl1996

Query Match 42.6%; Score 43; DB 2; Length 352;
Best Local Similarity 71.4%; Pred. No. 39;
Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 RAVRLALSGFDGRV 16
 ||| ||| | | |
Db 119 RAVLLALGGADGHV 132

RESULT 15
AC1310
chitinases homolog lmo1883 [imported] - *Listeria monocytogenes* (strain EGD-e)
C;Species: *Listeria monocytogenes*
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C;Accession: AC1310
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duseurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Maier,
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vasquez-Boland, J.A.; Voss, H.; Wehlend,
A;Title: Comparative genomics of *Listeria* species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AC1310
A;Status: preliminary
A:Molecule type: DNA
A;Residues: 1-352 <GLA>
A;Cross-references: GB:NC_003210; PIDN:CAC99961.1; PID:g16411336; GSPDB:GN00177
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: lmo1883

Query Match 42.6%; Score 43; DB 2; Length 352;
Best Local Similarity 71.4%; Pred. No. 39;
Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 RAVRLALSGFDGRV 16
 ||| ||| | | |
Db 119 RAVLLALGGADGHV 132

RESULT 16
AB3120
conserved hypothetical protein PA4204 [imported] - *Pseudomonas aeruginosa* (strain PAO1)
C;Species: *Pseudomonas aeruginosa*
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: AB3120
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bradman,
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen.
A;Reference number: AB32950; MUID:20437337; PMID:10984043
A;Accession: AB3120
A;Status: preliminary
A:Molecule type: DNA
A;Residues: 1-388 <STO>
A;Cross-references: GB:AE004837; GB:AE004091; NID:q9950414; PIDN:AAG07591.1; GSPDB:GN00178

```

RESULT 19
E99059
  A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
  A:Reference number: AD3252; PMID:11756688
  A:Accession: AH3290
  A:Status: preliminary
  A:Molecule type: DNA
  A:Residues: 1-335 <KUR>
  A:Cross-references: GB:AE008917; PIDN:AAL51491.1; PID:g17982204; GSPDB:GN00190
  A:Experimental source: strain 16M
  C:Genetics:
  A:Gene: BME10310
  A:Map position: I
  C:Superfamily: glyceraldehyde-3-phosphate dehydrogenase
  C:Keywords: oxidoreductase

  Query Match      42.1%; Score 42.5; DB 2; Length 335;
  Best Local Similarity 64.3%; Pred. No. 45;
  Matches 9; Conservative 4; Mismatches 0; Indels 1; Gaps 1;

  QY 4 AVRLLALSGFDGRVG 17
  Db 2 AVRVAINGF-GRIG 14
  |||:|:| |||:|
  |||:|:| |||:|

RESULT 22
AC0113
  D-erythrose 4-phosphate dehydrogenase (EC 1.2.1.-) [imported] - Yersinia pestis (strain
  C:Species: Yersinia pestis
  C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001
  C:Accession: AC0113
  R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
  deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
  il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
  Nature 413, 523-527, 2001
  A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
  A:Reference number: AB0001; MUID:21470413; PMID:11586360
  A:Accession: AC0113
  A:Status: preliminary
  A:Molecule type: DNA
  A:Residues: 1-338 <KUR>
  A:Cross-references: GB:AL590842; PIDN:CAC89766.1; PID:g15978993; GSPDB:GN00175
  C:Genetics:
  C:Superfamily: glyceraldehyde-3-phosphate dehydrogenase
  C:Keywords: oxidoreductase

  Query Match      42.1%; Score 42.5; DB 2; Length 338;
  Best Local Similarity 57.1%; Pred. No. 45;
  Matches 8; Conservative 5; Mismatches 0; Indels 1; Gaps 1;

  QY 4 AVRLLALSGFDGRVG 17
  Db 2 AIRIAINGF-GRIG 14
  |||:|:| |||:|
  |||:|:| |||:|

RESULT 23
AB2983
  conserved hypothetical protein Atu3467 [imported] - Agrobacterium tumefaciens (strain C58
  C:Species: Agrobacterium tumefaciens
  C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
  C:Accession: AB2983
  R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.;
  erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan
  ; Karp, P.; Romero, P.; Zhang, S.
  Science 294, 2317-2323, 2001
  A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
  ster, E.W.
  A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
  A:Reference number: AB2577; MUID:21608550; PMID:11743193
  A:Accession: AB2983
  A:Status: preliminary
  A:Molecule type: DNA
  A:Residues: 1-135 <KUR>
  A:Cross-references: GB:AE008689; PIDN:AAL44280.1; PID:g17741866; GSPDB:GN00187

```

A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu3467
A;Map position: linear chromosome

Query Match 41.6%; Score 42; DB 2; Length 135;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 9 LSGFGDGRVSGC 20

Db 86 LVGFGRMGSGC 97

RESULT 24

E98300
hypothetical protein AGR_L_2724 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C;Species: Agrobacterium tumefaciens
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C;Accession: E98300
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorllo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; MUID:21608551; PMID:11743194
A;Accession: E98300
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-135 <KUR>
A;Cross-references: GB:AE007870; PIDN:AAK89927.1; PID:g15159880; GSPDB:GN00170
C;Genetics:
A;Gene: AGR_L_2724
A;Map position: linear chromosome

Query Match 41.6%; Score 42; DB 2; Length 135;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 9 LSGFGDGRVSGC 20

Db 86 LVGFGRMGSGC 97

RESULT 25

AC1685
dihydrodipicolinate reductase homolog dapB [imported] - Listeria innocua (strain Clp112
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C;Accession: AC1685
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Psihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlant,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AC1685
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-263 <GLA>
A;Cross-references: GB:AL592022; PIDN:CAC97251.1; PID:g16414522; GSPDB:GN00178
A;Experimental source: strain Clp11262
C;Genetics:
A;Gene: dapB
C;Superfamily: dihydrodipicolinate reductase

Query Match 41.6%; Score 42; DB 2; Length 263;
Best Local Similarity 53.8%; Pred. No. 42;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 5 VRALSGFDGRVG 17

:::|||||

Db 1 MKVAUSGFKGRMG 13

RESULT 26

DCCHO

ornithine decarboxylase (EC 4.1.1.17) - chicken (fragment)

C;Species: Gallus gallus (Chicken)

C;Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 18-Jun-1999

C;Accession: A48386; S19892

R;Johnson, R.; Bulfield, G.

Anim. Genet. 23, 403-409, 1992

A;Title: Molecular cloning and sequence analysis of a chicken ornithine decarboxylase cD

A;Reference number: A48386; MUID:93036582; PMID:1416246

A;Accession: A48386

A;Molecule type: mRNA

A;Residues: 1-450 <JO2>

A;Cross-references: EMBL:X64710; NID:g63712; PIDN:CAA45965.1; PID:g63713

A;Experimental source: embryos

A;Note: sequence extracted from NCBI backbone (NCBIP:115105)

C;Function:

A;Description: This enzyme catalyzes the decarboxylation of ornithine to putrescine, the
C;Superfamily: ornithine decarboxylase
C;Keywords: carbon-carbon lyase; carboxy-lyase; phosphoprotein; polyamine biosynthesis; l
F;59/binding site: pyridoxal phosphate (Lys) (covalent) #status Predicted
F;350/Active site: Cys (shared with dimeric partner) #status Predicted

Query Match 41.6%; Score 42; DB 1; Length 450;
Best Local Similarity 50.0%; Pred. No. 73;
Matches 11; Conservative 3; Mismatches 6; Indels 2; Gaps 1;

QY 1 LORA--VRALSGFDGRVSGC 20

Db 171 LERAKELDLAIVGSFHVSGC 192

RESULT 27

T45585

hypothetical protein Fl1C1.230 - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000

C;Accession: T45585

R;Bargues, M.; Collado, M.C.; Navarro, P.; Terol, J.; Perez-Alonso, M.; Mewes, H.W.; Maye

submitted to the Protein Sequence Database, December 1999

A;Reference number: Z23007

A;Accession: T45585

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-466 <BAR>

A;Cross-references: EMBL:AL132976

A;Experimental source: cultivar Columbia; BAC clone Fl1C1

C;Genetics:

A;Map position: 3

A;Note: Fl1C1.230

Query Match 41.6%; Score 42; DB 2; Length 466;
Best Local Similarity 56.2%; Pred. No. 75;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 AVRLLSGFDGRVSGC 19

Db 252 AVNAVVSFGDGLVFTG 267

RESULT 28

A95926

probable argininosuccinate lyase (EC 4.3.2.1) [imported] - Sinorhizobium meliloti (strain

C;Species: Sinorhizobium meliloti

C;Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 14-Sep-2001

C;Accession: A95926

R;Finan, T.M.; Weidner, S.; Wong, K.; Ruhmester, J.; Chain, P.; Vorholter, P.J.; Hernan

Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001

A;Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing endo

A;Reference number: A95842; MUID:21396508; PMID:11481431

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2004, 06:49:47 ; Search time 7.66667 Seconds
(without alignments)
135.835 Million cell updates/sec

Title: 09549186-8

Perfect score: 101

Sequence: 1 LQRAVRLALSGFDGRVSGSC 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	44.6	263	1	DABP_LISMO
2	45	44.6	367	1	NOLF_RHIME
3	44	43.6	273	1	DABP_METTH
4	44	43.6	485	1	DPG2_HUMAN
5	44	43.6	3093	1	POLG_BSTV1
6	43.5	43.1	595	1	GLD2_MYCTU
7	43.5	43.1	693	1	RECG_HAEIN
8	43	42.6	393	1	DCO2_DROME
9	42.5	42.1	301	1	OTC_METTH
10	42	41.6	263	1	DABP_LISIN
11	42	41.6	394	1	DCOI_DROME
12	42	41.6	450	1	DCOR_CHICK
13	42	41.6	488	1	ARL2_RHIME
14	42	41.6	943	1	LBPA_NEIMB
15	42	41.6	944	1	LBPA_NEIMA
16	42	41.6	1872	1	T2D1_HUMAN
17	41.5	41.1	471	1	YEGB_ECOLI
18	41	40.6	251	1	YMP3_STRCO
19	41	40.6	255	1	DAPE_STRMU
20	41	40.6	259	1	DAPE_ENTFA
21	41	40.6	282	1	ROC_XENLA
22	40.5	40.1	334	1	G3P_PSEAE
23	40.5	40.1	374	1	G3PA_CHLRE
24	40.5	40.1	1078	1	CARB_BUCAL
25	40	39.6	229	1	MODE_ECOLI
26	40	39.6	257	1	DABP_CAUCR
27	40	39.6	279	1	YQA6_STRCO
28	40	39.6	355	1	MOAA_AERPE
29	40	39.6	431	1	DCOR_DATST
30	40	39.6	448	1	SPS2_HUMAN
31	40	39.6	1026	1	TAC2_HUMAN
32	39.5	39.6	1302	1	ACSC_ACEXY
33	39.5	39.1	331	1	G3P2_KLUMA

34	39.5	39.1	336	1	G3P1_STAAM
35	39.5	39.1	337	1	G3P_MYCGE
36	39.5	39.1	337	1	G3P_MYCPN
37	39.5	39.1	337	1	G3P_ZYMMO
38	39.5	39.1	986	1	YHDP_ECOLI
39	39	38.6	256	1	YDHI_HSVSC
40	39	38.6	275	1	RK2_PICAB
41	39	38.6	276	1	RR2_PINTH
42	39	38.6	285	1	YX47_STRCO
43	39	38.6	293	1	NANA_STAAM
44	39	38.6	293	1	NANA_STAAM
45	39	38.6	293	1	YF75_MYCTU
46	39	38.6	306	1	ROC_HUMAN
47	39	38.6	313	1	ROC_MOUSE
48	39	38.6	426	1	CSIE_ECOLI
49	39	38.6	426	1	SYS_ANASP
50	39	38.6	437	1	RECA_TROWT
51	39	38.6	490	1	TWST_DROME
52	39	38.6	514	1	ATPA_THIFE
53	39	38.6	546	1	DX56_MOUSE
54	39	38.6	547	1	DX56_HUMAN
55	39	38.6	570	1	TREA_SALTI
56	39	38.6	570	1	TREA_SALTY
57	39	38.6	592	1	VATA_METKA
58	39	38.6	611	1	IF4B_HUMAN
59	39	38.6	664	1	SYM_BACSU
60	39	38.6	895	1	POL_RSVF
61	39	38.6	1477	1	YORI_YEAST
62	39	38.6	2514	1	POLN_ONNVG
63	38.5	38.1	292	1	ARBL_DEIRA
64	38.5	38.1	331	1	G3P1_YEAST
65	38.5	38.1	331	1	G3P2_YEAST
66	38.5	38.1	331	1	G3P3_YEAST
67	38.5	38.1	336	1	G3P1_STAEP
68	38.5	38.1	338	1	E4PD_ECOLI
69	38.5	38.1	414	1	G3PA_CHOCR
70	38.5	38.1	416	1	G3PA_GRAVE
71	38.5	38.1	514	1	IMD1_HUMAN
72	38.5	38.1	514	1	IMD1_MOUSE
73	38.5	38.1	557	1	ILVD_AERPE
74	38.5	38.1	598	1	RACA_DICDI
75	38	37.6	104	1	RS10_SYNEL
76	38	37.6	127	1	YHHK_ECOLI
77	38	37.6	132	1	CHAZ_BOMMO
78	38	37.6	146	1	CDD_MOUSE
79	38	37.6	154	1	YK01_AERPE
80	38	37.6	180	1	NUSG_ECOL6
81	38	37.6	180	1	NUSG_ECOLI
82	38	37.6	183	1	NUSG_PASMU
83	38	37.6	214	1	NIRD_PARDE
84	38	37.6	243	1	PYRF_XANAC
85	38	37.6	243	1	PYRF_XANCP
86	38	37.6	250	1	DABP_STRAP
87	38	37.6	255	1	DAEP_STRPN
88	38	37.6	261	1	DABP_LACPL
89	38	37.6	269	1	DABP_VIBCH
90	38	37.6	269	1	DABP_VIBPA
91	38	37.6	270	1	RECK_LACLA
92	38	37.6	274	1	DABP_CANBF
93	38	37.6	331	1	CSCR_ECOLI
94	38	37.6	341	1	RTCA_PSEAE
95	38	37.6	353	1	DCAM_PEA
96	38	37.6	353	1	DCAM_VICFA
97	38	37.6	366	1	QUEA_CAUCR
98	38	37.6	459	1	EX7L_YERPE
99	38	37.6	482	1	PRPD_ECO57
100	38	37.6	482	1	PRPD_ECOLI

ALIGNMENTS

RESULT 1

DABP_LISMO
 ID DABP_LISMO STANDARD; PRT; 263 AA.
 AC Q8Y526;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DT Dihydrodipicolinate reductase (EC 1.3.1.26) (DHPR).
 DE DABP OR LMO1907.
 OS Listeria monocytogenes.
 OS Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
 OX NCBI_TaxID=1639;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EGD-e / Serovar 1/2a;
 RC MEDLINE=21537279; PubMed=11679669;
 RA Glaeser P., Frangeul L., Buchrieser C., Ruoniok C., Amend A.,
 RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
 RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehroux P.,
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussauguet O.,
 RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
 RA Jones L.-M., Kaerst U., Krefz J., Kuhn M., Kunst F., Kurapkat G.,
 RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
 RA Nordisiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
 RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
 RA Vazquez-Boland J.-A., Voss H., Weiland J., Cossart P.;
 RT "Comparative genomics of *Listeria* species";
 RL Science 294:849-852(2001).
 CC -1- CATALYTIC ACTIVITY: 2,3,4,5-tetrahydrodipicolinate + NAD(P)(+) =
 CC 2,3-dihydrodipicolinate + NAD(P)H.
 CC -1- PATHWAY: Biosynthesis of diaminomelate and lysine from aspartate
 CC semialdehyde; second step.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: Belongs to the dihydrodipicolinate reductase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>
 CC or send an email to license@ebi.ac.uk).
 CC
 CC EMBL; AL591981; CAC99985.1; -;
 CC FIC; AC1313; AC1313.
 CC List18c; LMO01907; -;
 CC HAMAP; MF_00102; -; 1.
 CC InterPro; IPR000846; DabP.
 CC Pfam; PF05173; DabP_C; 1.
 CC Pfam; PF01113; DabP_N; 1.
 CC ProDom; PD004105; DabP; 1.
 CC TIGRfam; TIGR00036; dabP; 1.
 CC PROSITE; PS01298; DABP; FALSE NEG.
 KW Diaminomelate biosynthesis; Lysine biosynthesis; Oxidoreductase;
 KW NADP; Complete proteome.
 SQ SEQUENCE 263 AA; 28900 MW; C138247A0426DF0C CRC64;
 Query Match 44.6%; Score 45; DB 1; Length 263;
 Best Local Similarity 61.5%; Pred. No. 6.5;
 Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
 QY 5 VRLALSGFDGRVG 17
 DB 1 MRVAVSGFKGRWG 13
 RESULT 2
 NOLF RHIME STANDARD; PRT; 367 AA.
 ID NOLF RHIME
 AC P25136;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Dihydrodipicolinate reductase (EC 1.3.1.26) (DHPR).
 OS Listeria monocytogenes.
 OS Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
 OX NCBI_TaxID=1639;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EGD-e / Serovar 1/2a;
 RC MEDLINE=21537279; PubMed=11679669;
 RA Glaeser P., Frangeul L., Buchrieser C., Ruoniok C., Amend A.,
 RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
 RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehroux P.,
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussauguet O.,
 RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
 RA Jones L.-M., Kaerst U., Krefz J., Kuhn M., Kunst F., Kurapkat G.,
 RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
 RA Nordisiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
 RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
 RA Vazquez-Boland J.-A., Voss H., Weiland J., Cossart P.;
 RT "Comparative genomics of *Listeria* species";
 RL Science 294:849-852(2001).
 CC -1- CATALYTIC ACTIVITY: 2,3,4,5-tetrahydrodipicolinate + NAD(P)(+) =
 CC 2,3-dihydrodipicolinate + NAD(P)H.
 CC -1- PATHWAY: Biosynthesis of diaminomelate and lysine from aspartate
 CC semialdehyde; second step.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: Belongs to the dihydrodipicolinate reductase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>
 CC or send an email to license@ebi.ac.uk).
 CC
 CC EMBL; AL591981; CAC99985.1; -;
 CC FIC; AC1313; AC1313.
 CC List18c; LMO01907; -;
 CC HAMAP; MF_00102; -; 1.
 CC InterPro; IPR000846; DabP.
 CC Pfam; PF05173; DabP_C; 1.
 CC Pfam; PF01113; DabP_N; 1.
 CC ProDom; PD004105; DabP; 1.
 CC TIGRfam; TIGR00036; dabP; 1.
 CC PROSITE; PS01298; DABP; FALSE NEG.
 KW Diaminomelate biosynthesis; Lysine biosynthesis; Oxidoreductase;
 KW NADP; Complete proteome.
 SQ SEQUENCE 263 AA; 28900 MW; C138247A0426DF0C CRC64;
 Query Match 44.6%; Score 45; DB 1; Length 263;
 Best Local Similarity 61.5%; Pred. No. 6.5;
 Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
 QY 5 VRLALSGFDGRVG 17
 DB 1 MRVAVSGFKGRWG 13
 RESULT 2
 NOLF RHIME STANDARD; PRT; 367 AA.
 ID NOLF RHIME
 AC P25136;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Dihydrodipicolinate reductase (EC 1.3.1.26) (DHPR).
 OS Listeria monocytogenes.
 OS Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
 OX NCBI_TaxID=1639;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EGD-e / Serovar 1/2a;
 RC MEDLINE=21537279; PubMed=11679669;
 RA Glaeser P., Frangeul L., Buchrieser C., Ruoniok C., Amend A.,
 RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
 RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehroux P.,
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussauguet O.,
 RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
 RA Jones L.-M., Kaerst U., Krefz J., Kuhn M., Kunst F., Kurapkat G.,
 RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
 RA Nordisiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
 RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
 RA Vazquez-Boland J.-A., Voss H., Weiland J., Cossart P.;
 RT "Comparative genomics of *Listeria* species";
 RL Science 294:849-852(2001).
 CC -1- CATALYTIC ACTIVITY: 2,3,4,5-tetrahydrodipicolinate + NAD(P)(+) =
 CC 2,3-dihydrodipicolinate + NAD(P)H.
 CC -1- PATHWAY: Biosynthesis of diaminomelate and lysine from aspartate
 CC semialdehyde; second step.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: Belongs to the dihydrodipicolinate reductase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>
 CC or send an email to license@ebi.ac.uk).
 CC
 CC EMBL; AL591981; CAC99985.1; -;
 CC FIC; AC1313; AC1313.
 CC List18c; LMO01907; -;
 CC HAMAP; MF_00102; -; 1.
 CC InterPro; IPR000846; DabP.
 CC Pfam; PF05173; DabP_C; 1.
 CC Pfam; PF01113; DabP_N; 1.
 CC ProDom; PD004105; DabP; 1.
 CC TIGRfam; TIGR00036; dabP; 1.
 CC PROSITE; PS01298; DABP; FALSE NEG.
 KW Diaminomelate biosynthesis; Lysine biosynthesis; Oxidoreductase;
 KW NADP; Complete proteome.
 SQ SEQUENCE 263 AA; 28900 MW; C138247A0426DF0C CRC64;
 Query Match 44.


```
SQ SEQUENCE 485 AA; 54911 MW; B99734BFEA249192 CRC64;
Query Match 43.6%; Score 44; DB 1; Length 485;
Best Local Similarity 52.9%; Pred. No. 18;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 3 RAVRLALSGDGRVSGG 19
DQ 12 KVCRCLLSGFGRVDAG 28

RESULT 5
POLG BSTV1
ID POLG BSTV1 STANDARD; PRT; 3093 AA.
AC Q65730;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polyprotein [Contains: N-terminal protein (P1); Helper
DE protein 1 (6K1); Cytoplasmic inclusion protein (CI); 6 kDa protein 2
DE (6K2); Genome-linked protein (VPG); Nuclear inclusion protein A (NI-A)
DE (NIA) (EC 3.4.22.44) (49 kDa proteinase) (49 kDa-Pro); Nuclear
DE inclusion protein B (NI-B) (NIB) (RNA-directed RNA polymerase)
DE (EC 2.7.7.48); Coat protein (CP)].
OS Brome streak virus (strain 11-Cal) (BstV) (Brome streak mosaic
OS rymovirus).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potyviridae;
OC Tritimovirus.
OX NCBI_TaxID=117138;
RN [1] _SEQUENCE FROM N.A.
RX MEDLINE=95363423; PubMed=7636484;
RA Goetz R., Maiss E.;
RT "The complete nucleotide sequence and genome organization of the
RT mite-transmitted brome streak mosaic rymovirus in comparison with
RT those of potyviruses.";
RL J. Gen. Virol. 76:2035-2035 (1995).
CC -!- FUNCTION: Helper component-proteinase is required for aphid
CC transmission and also has proteolytic activity.
CC -!- FUNCTION: Cytoplasmic inclusion protein has helicase activity. It
CC may be involved in replication.
CC -!- FUNCTION: Nuclear inclusion protein A has proteolytic activity.
CC -!- CATALYTIC ACTIVITY: Hydrolyzes glutamyl bonds, and activity is
CC further restricted by preferences for the amino acids in P6 - P1'
CC that vary with the species of potyvirus, e.g. Glu-Xaa-Xaa-Tyr-Xaa-
CC Glu-(Ser or Gly) for the enzyme from tobacco etch virus. The
CC natural substrate is the viral polyprotein, but other proteins and
CC oligopeptides containing the appropriate consensus sequence are
CC also cleaved.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC -!- CATALYTIC ACTIVITY: Hydrolyzes a Gly-|-Gly bond at its own C-
CC terminus, commonly in the sequence -Tyr-Xaa-Val-Gly-|-Gly, in the
CC processing of the potyviral polyprotein.
CC -!- PTM: VPG is covalently linked to the genomic RNA.
CC -!- PTM: The viral RNA of potyviruses is expressed as a single
CC polyprotein which undergoes posttranslational proteolytic
CC processing resulting in the production of at least eight
CC individual proteins.
CC -!- SIMILARITY: HC PROTEINASE BELONGS TO PEPTIDASE FAMILY C6.
CC -!- SIMILARITY: NI-A PROTEINASE BELONGS TO PEPTIDASE FAMILY C4.
CC -!- SIMILARITY: Belongs to the potyviruses polyprotein family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z48506; CAA88417.1; -.
```

```
DR MEROPS; C04.UPW; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR001730; Peptidase_C4.
DR InterPro; IPR001592; Poty_coat.
DR InterPro; IPR007095; RNA_pol_PS.
DR InterPro; IPR001205; RNA_pol_P3D.
DR InterPro; IPR007094; RNA_pol_P5vir.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00863; Peptidase_C4; 1.
DR Pfam; PF00767; Poty_coat; 1.
DR Pfam; PF00680; RNA_dep_RNA_pol; 1.
DR PRINTS; PR00966; NIAPOTYPTRASE.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 1.
KW Hydrolase; Transferase; Thiol protease; RNA-directed RNA polymerase;
KW Coat protein; Polyprotein; Covalent protein-RNA linkage; Helicase;
KW ATP-binding.
FT CHAIN 1 403 N-TERMINAL PROTEIN.
FT CHAIN 404 809 HELPER COMPONENT PROTEINASE.
FT CHAIN 810 1087 PROTEIN P3.
FT CHAIN 1088 1138 6 kDa protein 1.
FT CHAIN 1139 1783 CYTOPLASMIC INCLUSION PROTEIN.
FT CHAIN 1784 1834 6 kDa protein 2.
FT CHAIN 1835 2040 GENOME-LINKED PROTEIN.
FT CHAIN 2041 2275 NUCLEAR INCLUSION PROTEIN A.
FT CHAIN 2276 2827 NUCLEAR INCLUSION PROTEIN B.
FT CHAIN 2828 3093 COAT PROTEIN.
FT BINDING 1915 1915 COVALENT LINKAGE OF VIRAL RNA
FT (BY SIMILARITY).
FT NP_BIND 1228 1235 ATP (POTENTIAL).
FT SEQUENCE 3093 AA; 348101 MW; 50F775CE7ABCCB4 CRC64;
Query Match 43.6%; Score 44; DB 1; Length 3093;
Best Local Similarity 47.1%; Pred. No. 1.2e+02;
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 4 AVRRLALSGDGRVSGG 20
DQ 981 AYRLTYTGVARIGRC 997

RESULT 6
GLD2_MYCTU
ID GLD2_MYCTU STANDARD; PRT; 585 AA.
AC Q07168;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Glycerol-3-phosphate dehydrogenase 2 (EC 1.1.99.5).
GN GLP22 OR RV3302C OR MT3401 OR MTC1418A.04C OR MTV016.01C OR MB3330C.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1] _SEQUENCE FROM N.A.
RP SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmler K., Gas S., Barry C.E. III, Tekaiia F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
```


SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
 MEDLINE=22206494; PubMed=12218036;
 Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
 "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 laboratory strains.";
 J. Bacteriol. 184:5479-5490(2002).
 [3]
 SEQUENCE FROM N.A.
 SPECIES=M.bovis; STRAIN=AF2122/97;
 MEDLINE=22709107; PubMed=12788972;
 Garner T., Eiglmeyer K., Camus J.-C., Lacroix C., Monsemp C., Simon S.,
 Pryor M., Duthoy S., Gronin S., Dognett J., Mayes R., Keating L., Wheeler P.R.,
 Harris B., Atkin R., Goggin S., Dognett J., Mayes R., Keating L., Wheeler P.R.,
 Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
 "The complete genome sequence of Mycobacterium bovis.";
 Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 -/- CATALYTIC ACTIVITY: Sn-glycerol 3-phosphate + acceptor = glycerone
 phosphate + reduced acceptor.
 -/- COFACTOR: FAD (By similarity).
 -/- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 -/- SIMILARITY: Belongs to the FAD-dependent glycerol-3-phosphate
 dehydrogenase family.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).

 EMBL: Z36070; CAB09447.1; -;
 EMBL: AE007149; AAK47744.1; -;
 EMBL: BX248345; CAD95422.1; -;
 EMBL: H70533; H70533.
 TIGR: MT3401; -;
 Tuberculist; RV3302C; -;
 InterPro: IPR000447; FAD_Gly3P_dh.
 InterPro: IPR006076; FAD_Oxred.
 InterPro: IPR000205; NAD_BS.
 InterPro: IPR001100; Pyr_redox.
 Pfam: PF01266; DAO; 1.
 PRINTS: PR01001; FADG3PDH.
 PRINTS: PR00411; FNDREDTASE1.
 PROSITE: PS00977; FAD_G3PDH_1; 1.
 PROSITE: PS00978; FAD_G3PDH_2; 1.
 Glycylol metabolism; Oxidoreductase; Flavoprotein; FAD;
 Complete proteome.
 NP_BIND 37 FAD (BY SIMILARITY).
 SQ SEQUENCE 585 AA; 62777 MW; 760C15F7E75BB69F CRC64;
 Query Match 43.1%; Score 43.5; DB 1; Length 585;
 Best Local Similarity 50.0%; Pred. No. 26;
 Matches 14; Conservative 0; Mismatches 5; Indels 9; Gaps 2;
 QY 2 QRAV---RLALSGFD-----GRVSGGC 20
 ||||| ||||| ||||| ||||| |||||
 23 QRAVAKRLGTEQFDVVWIGGGVSGGC 50
 Db
 RESULT 7
 RECG HAEIN STANDARD; PRT; 693 AA.
 ID RECG HAEIN STANDARD; PRT; 693 AA.
 AC P43809;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE ATP-dependent DNA helicase recG (EC 3.6.1.-).
 GN RECG OR H11740.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=rd / KW20 / ATCC 51907;
 RX MEDLINE=95350830; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Doughterty B.A., Merrick J.M.,
 Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 Fine L.D., Fritchman J.L., Fuhrmann J.L., Saudek D.M., Brandon R.C.,
 Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 Venter J.C.;
 "Whole-genome random sequencing and assembly of Haemophilus influenzae
 Rd.";
 Science 269:496-512(1995).
 -/- FUNCTION: Critical role in recombination and DNA repair. Help
 process Holliday junction intermediates to mature products by
 catalyzing branch migration. Has a DNA unwinding activity
 characteristic of a DNA helicase with a 3' to 5' polarity. RecG
 unwind branched duplex DNA (Y-DNA) (By similarity).
 -/- SIMILARITY: Belongs to the helicase family. RecG subfamily.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).

 EMBL: U32847; AAC23387.1; -;
 EMBL: E64139; B64139.
 TIGR: H11740; -;
 InterPro: IPR001410; DEAD.
 InterPro: IPR001650; Helicase_C.
 InterPro: IPR008994; Nucleic_acid_OB.
 InterPro: IPR004609; RecG.
 InterPro: IPR004365; tRNA_ant1.
 Pfam: PF00270; DEAD; 1.
 Pfam: PF00271; helicase_C; 1.
 Pfam: PF01336; tRNA_ant1; 1.
 SMART: SM00487; DEXDC; 1.
 SMART: SM00490; HELICCC; 1.
 TIGRFAMs: TIGR00643; recG; 1.
 Hydrolase; Helicase; DNA repair; ATP-binding; DNA recombination;
 DNA-binding; Complete proteome.
 NP_BIND 296 303 ATP (POTENTIAL).
 SITE 397 400 DEQH BOX.
 SQ SEQUENCE 693 AA; 78085 MW; 12D1BC65F0259950 CRC64;
 Query Match 43.1%; Score 43.5; DB 1; Length 693;
 Best Local Similarity 47.8%; Pred. No. 31;
 Matches 11; Conservative 2; Mismatches 7; Indels 3; Gaps 1;
 QY 1 LQAVRLALSGFD---GRVSGGC 20
 ::|||::|||::|||::|||::|||
 567 IENAVRLGLSQLHQLRGRVGRGC 589
 Db
 RESULT 8
 DCO2_DROME STANDARD; PRT; 393 AA.
 ID DCO2_DROME STANDARD; PRT; 393 AA.
 AC P40808;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Ornithine decarboxylase 2 (EC 4.1.1.17) (ODC).
 GN ODC2.


```

FT CONFLICT 44 45 NV -> KL (IN REF. 2).
FT CONFLICT 242 242 K -> Q (IN REF. 1; CAA47165).
FT CONFLICT 325 325 L -> Q (IN REF. 1; CAA47165).
FT CONFLICT 348 348 K -> Q (IN REF. 1; CAA47165).
SQ SEQUENCE 394 AA; 44166 MW; CD90520632B7D3B2 CRC64;

Query Match 41.6%; Score 42; DB 1; Length 394;
Best Local Similarity 40.0%; Pred. No. 30;
Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 LQRAVRLALSGFDGRVSGC 20
Db 177 LAKSLKVTGTSTFHVSGC 196

RESULT 12
DCOR CHICK STANDARD; PRT; 450 AA.
AC P27118;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ornithine decarboxylase (EC 4.1.1.17) (ODC) (Fragment).
GN ODC.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=White Leghorn;
RX MEDLINE=93036582; PubMed=1416246;
RA Johnson R., Bulfield G.;
RT "Molecular cloning and sequence analysis of a chicken ornithine
RT decarboxylase cDNA.";
RL Anim. Genet. 23:403-409(1992).
CC -1- CATALYTIC ACTIVITY: L-ornithine = putrescine + CO(2).
CC -1- COFACTOR: Pyridoxal phosphate.
CC -1- PATHWAY: Polyamine biosynthesis, first (rate-limiting) step.
CC -1- SUBUNIT: Homodimer.
CC -1- SIMILARITY: BELONGS TO FAMILY 2 OF ORNITHINE, DAP, AND ARGinine
CC DECARBOXYLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X64710; CAA45965.1; -.
CC PIR; A48386; DCCHO.
CC HSSP; P11926; 1D7K.
CC InterPro; IPR000183; Decarboxylase2.
CC DR InterPro; IPR009006; Racem_decarbox_C.
CC DR Pfam; PF02784; Orn_Arg_Gec_N; 1.
CC DR Pfam; PF00278; Orn_DAP_Arg_Gec; 1.
CC DR PRINTS; PR01179; ODACRBLXASE.
CC DR PROSITE; PS00878; ODR_DC_2_1; 1.
CC DR PROSITE; PS00879; ODR_DC_2_2; 1.
CC Lyase; Decarboxylase; Pyridoxal phosphate; Polyamine biosynthesis;
KW Phosphorylation.
FT NON_TER 1
FT BINDING 59 59 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT ACT_SITE 350 350 BY SIMILARITY.
FT MOD_RES 293 293 PHOSPHORYLATION (BY CK2)
FT (BY SIMILARITY).
SQ SEQUENCE 450 AA; 49734 MW; 5EAC75DF3D17AD2C CRC64;

Query Match 41.6%; Score 42; DB 1; Length 450;
Best Local Similarity 50.0%; Pred. No. 35;
Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 1 LQRAVRLALSGFDGRVSGC 20
Db 177 LAKSLKVTGTSTFHVSGC 196

RESULT 13
ARL2 RHIME
ID ARL2 RHIME STANDARD; PRT; 488 AA.
AC Q92VM6;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Argininosuccinate lyase 2 (EC 4.3.2.1) (Argininosuccinase 2) (ASAL 2).
GN ARGH2 OR RB0673 OR SMB21094.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OG Plasmid pSymB (megaplasmid 2).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396508; PubMed=11481431;
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
RA Vorholter F.J., Hernandez-Lucas I., Becker A., Gouzy J.,
RA Golding B., Puehler A.;
RT "The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-
RT fixing endosymbiont Sinorhizobium meliloti.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
CC -1- CATALYTIC ACTIVITY: N-(L-arginino)succinate = fumarate + L-
CC arginine.
CC -1- PATHWAY: Arginine biosynthesis; eighth (last) step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: Belongs to the lyase 1 family. Argininosuccinate lyase
CC subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL603644; CAC49073.1; -.
CC PIR; A95926; A95926.
CC HAMAP; MF_00006; -.
CC InterPro; IPR009049; argH.
CC DR InterPro; IPR000362; Fumarate_lyase.
CC DR InterPro; IPR008948; L-Aspartase-like.
CC DR Pfam; PF0206; lyase_1; 1.
CC DR PRINTS; PR00149; FUMRATELYASE.
CC DR TIGRFAMs; TIGR00838; argH; 1.
CC DR PROSITE; PS00163; FUMARATE_LYASES; 1.
CC Arginine biosynthesis; lyase; Plasmid; Complete proteome.
KW Arginine biosynthesis; lyase; Plasmid; Complete proteome.
SQ SEQUENCE 488 AA; 53037 MW; 376898B050516832 CRC64;

Query Match 41.6%; Score 42; DB 1; Length 488;
Best Local Similarity 50.0%; Pred. No. 38;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 4 AVRLLALSGFDGRVSGC 19
Db 61 AIRAALRGIEADVAG 76

RESULT 14
LEPA NEIMB
ID LEPA_NEIMB STANDARD; PRT; 943 AA.
AC Q06379; Q0JYK5;
DT 01-NOV-1995 (Rel. 32, Created)

```

DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Lactoferrin binding protein A precursor (Iron-regulated outer membrane protein A).
 DE LBP A OR IROA OR NMB1540.
 GN Neisseria meningitidis (serogroup B).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 OC Neisseriaceae; Neisseria.
 OX NCBI_TaxID=491;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=MC58 / Serogroup B;
 RX MEDLINE=20175755; PubMed=10710307;
 RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E., Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J., Nelson W.C., Gwin M.L., DeBoy R., Peterson J.D., Hickey E.K., Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A., Mason T., Ciecko A., Parksey D.S., Blair E., Citterio H., Clark E.B., Ratton M.D., Uterback T.R., Khouri H., Qin H., Vamathevan J., Gill J., Scarlato V., Masignani V., Pizzo M., Grandi G., Sun L., Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
 RT "Complete genome sequence of Neisseria meningitidis serogroup B strain MC58";
 RL Science 287:1809-1815 (2000).
 CC -!- FUNCTION: UNKNOWN. MAY BE AN IRON-SIDEROPHORE RECEPTOR.
 CC -!- SUBCELLULAR LOCATION: Outer membrane.
 CC -!- INDUCTION: By iron starvation.
 CC -!- SIMILARITY: Belongs to the tonB-dependent receptor family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X69214; CAA49148.1; --
 CC EMBL; AE002504; BAF41895.1; --
 CC FIC; G81070; G81070.
 CC TIGR; NMB1540; --
 DR InterPro; IPR000531; TonB boxC.
 DR Pfam; PF00593; TonB dep Rec; 1.
 DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
 DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
 KW Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
 KW Complete proteome.
 FT SIGNAL 1 27
 FT CHAIN 28 943
 FT SITE 826 943
 FT P-> Q (IN REF. 1).
 FT IAT -> VAA (IN REF. 1).
 FT A -> S (IN REF. 1).
 FT QAGAT -> NPETA (IN REF. 1).
 FT I -> V (IN REF. 1).
 FT V -> A (IN REF. 1).
 FT R -> H (IN REF. 1).
 FT E -> A (IN REF. 1).
 FT D -> N (IN REF. 1).
 FT DIKKTPPPFVS -> GIKKPEGGGYFLA (IN REF. 1).
 FT RES -> SEL (IN REF. 1).
 FT L -> V (IN REF. 1).
 FT YGK -> NGN (IN REF. 1).
 FT Q -> M (IN REF. 1).

FT CONFLICT 389 389 E -> K (IN REF. 1).
 FT CONFLICT 420 423 KNLV -> OKLI (IN REF. 1).
 FT CONFLICT 455 455 A -> K (IN REF. 1).
 FT CONFLICT 546 546 K -> N (IN REF. 1).
 FT CONFLICT 564 572 STGFENNO -> YSDYTDKG (IN REF. 1).
 FT CONFLICT 658 658 L -> V (IN REF. 1).
 FT CONFLICT 667 667 V -> L (IN REF. 1).
 SQ SEQUENCE 943 AA; 105680 MW; BD569ECACFC01A84 CRC64;
 Query Match 41.6%; Score 42; DB 1; Length 943;
 Best Local Similarity 60.0%; Pred. No. 74;
 Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 5 VRLASGDFGRVGS 19
 Db 126 VQSLSGVGRGSG 140
 RESULT 15
 LBPA NEIMA
 ID LBPA NEIMA STANDARD; PRT; 944 AA.
 AC Q9JTK4;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Lactoferrin binding protein A precursor.
 GN LBPA OR NMA1739.
 OS Neisseria meningitidis (serogroup A).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 OC Neisseriaceae; Neisseria.
 OX NCBI_TaxID=65699;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=Z2491 / Serogroup A / Serotype 4A;
 RX MEDLINE=20222556; PubMed=10761919;
 RA Parkhill J., Achtman G., James K.D., Bentley S.D., Churcher C., Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T., Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S., Jagels K., Leather S., Moule S., Mungall K., Quail M.A., Rajadream K.M., Rutherford K.M., Simmonds M., Skelton J., Whitehead S., Spratt B.G., Barrell B.G.;
 RA "Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491";
 RT Nature 404:502-506 (2000).
 RL Nature 404:502-506 (2000).
 CC -!- FUNCTION: UNKNOWN. MAY BE AN IRON-SIDEROPHORE RECEPTOR.
 CC -!- SUBCELLULAR LOCATION: Outer membrane.
 CC -!- SIMILARITY: Belongs to the tonB-dependent receptor family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AL162757; CAB84967.1; --
 CC FIC; C81798; C81798.
 CC InterPro; IPR000531; TonB boxC.
 CC Pfam; PF00593; TonB dep Rec; 1.
 CC PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
 CC PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
 KW Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
 KW Complete proteome.
 FT SIGNAL 1 27
 FT CHAIN 28 944
 FT SITE 827 944
 FT TONB C-TERMINAL BOX.
 SQ SEQUENCE 944 AA; 105732 MW; CE06B6192E74AE3E CRC64;
 Query Match 41.6%; Score 42; DB 1; Length 944;
 Best Local Similarity 60.0%; Pred. No. 74;
 Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

```
QY      5 VRLALSGFDGRVGS 19
      1  : ||| : |||
Db      126 VQSLSGYGRGSG 140

RESULT 16
T2D1 HUMAN
ID T2D1 HUMAN STANDARD; PRT; 1872 AA.
AC P21675;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Transcription initiation factor TFIID 250 kDa subunit (TAFII-250)
DE (TAFII250) (TBP-associated factor 250 kDa) (P250) (Cell cycle gene 1
DE protein).
DE TAFI1 OR TAF2A OR CCG1 OR BAP2.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Laryngeal carcinoma;
RX MEDLINE=91246200; PubMed=2038334;
RA Sekiguchi T., Nohiro Y., Nakamura Y., Hisamoto N., Nishimoto T.;
RT "The human CCG1 gene, essential for progression of the G1 phase,
RT encodes a 210-kilodalton nuclear DNA-binding protein.";
RL Mol. Cell. Biol. 11:3317-3325(1991).
RN [2]
RP PRELIMINARY SEQUENCE FROM N.A.
RX MEDLINE=89005056; PubMed=3169001;
RA Sekiguchi T., Miyata T., Nishimoto T.;
RT which complements the temperature-sensitive Gi mutants, tsBN462 and
RT ts13, of the BHK cell line.";
RL EMBO J. 7:1683-1687(1988).
RN [3]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=93196704; PubMed=7680771;
RA Ruppert S., Wang B.H., Tjian R.;
RT "Cloning and expression of human TAFII250: a TBP-associated factor
RT implicated in cell-cycle regulation.";
RL Nature 362:175-179(1993).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=93196705; PubMed=8450888;
RA Hisatake K., Hasegawa S., Takada R., Nakatani Y., Horikoshi M.,
RA Roeder R.G.;
RT "The p250 subunit of native TATA box-binding factor TFIID is the
RT cell-cycle regulatory protein CCG1.";
RL Nature 362:179-181(1993).
CC -!- FUNCTION: May play an essential role in TFIID assembly by
CC interacting with both TBP and other TAF, as well as serving to
CC link the control of transcription to the cell cycle. Essential
CC for progression of the G1 phase of the cell cycle. Possesses
CC DNA-binding activity.
CC -!- SUBUNIT: TFIID is composed of TBP and a variety of TBP-associated
CC factors.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- PTM: Phosphorylated by casein kinase II in vitro.
CC -!- SIMILARITY: Contains 2 bromodomains.
CC -!- SIMILARITY: Contains 1 HMG box domain.
CC -!- SIMILARITY: TO DROSOPHILA TAFII-230. SOME TO S.POMBE TAFII-111
CC AND TO S.CEREVISIAE TAF145.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

```
DR EMBL; D90359; BAA14374.1; -.
DR EMBL; X07024; CAA30073.1; ALT_SEQ.
DR PIR; A40262; A40262
DR PDB; 1EQF; 07-JUN-00.
DR TRANSFAC; T02206; -.
DR Genew; HGNC:11535; TAF1.
DR MIM; 313650; -.
DR GO; GO:0005669; C:transcription factor TFIID complex; TAS.
DR GO; GO:0016251; F:general RNA polymerase II transcription fac. .; TAS.
DR GO; GO:0004672; F:protein kinase activity; TAS.
DR GO; GO:0000114; F:Gi-specific transcription in mitotic cell c. .; TAS.
DR GO; GO:0006468; P:protein amino acid phosphorylation; TAS.
DR GO; GO:0006461; P:protein complex assembly; TAS.
DR InterPro; IPR001487; Bromodomain.
DR Pfam; PF00439; Bromodomain; 2.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 2.
DR PROSITE; PS00633; BROMODOMAIN 1; 2.
DR PROSITE; PS0014; BROMODOMAIN 2; 2.
DR Bromodomain; Nuclear protein; DNA-binding; Cell cycle; Repeat;
KW Transcription regulation; Phosphorylation; 3D-structure.
FT DOMAIN 157 165 PRO-RICH.
FT DNA BIND 1195 1273 HMG BOX (POTENTIAL).
FT DOMAIN 1351 1358 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 1397 1467 BROMODOMAIN 1.
FT DOMAIN 1520 1590 BROMODOMAIN 2.
FT DOMAIN 1627 1872 ASP/GLU-RICH (ACIDIC TAIL).
SQ SEQUENCE 1872 AA; 212676 MW; 93BE3D181A72ABEB CRC64;

Query Match 41.6%; Score 42; DB 1; Length 1872;
Best Local Similarity 56.2%; Pred. No. 1.5e+02;
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY      4 AVRLLALSGFDGRVGS 19
      1  : ||| : |||
Db      1740 AIQLSESGSDSDVGS 1755

RESULT 17
YEGB_ECOLI
ID YEGB_ECOLI STANDARD; PRT; 471 AA.
AC P36554; P76400;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical transport protein yegB.
GN YEGB OR B2077.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251358; PubMed=9097040;
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T.,
RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T.,
RA Saiki N., Sampei G., Seki Y., Sivasubram S., Tagami H.,
RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 40.1-50.0 min region on the linkage map.";
RN [2]
```

```

RL DNA Res. 3:379-392(1996).
RN [3]
RP SEQUENCE OF 381-471 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=94110256; PubMed=8282725;
RA Nagaawa S., Ishige K., Mizuno T.;
RT "Novel members of the two-component signal transduction genes in
EScherichia coli.";
RL J. Biochem. 114:350-357(1993).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Probable).
CC -!- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN
CC AS THE DRUG RESISTANCE TRANSLOCASE FAMILY).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; A5000297; AAC75138.1; -.
DR EMBL; D90846; BAA15933.1; -.
DR EMBL; D90847; BAA15944.1; -.
DR EMBL; D14054; BAA03139.1; -.
DR PIR; D64974; D64974.
DR EcoGene; EG12136; vegB.
DR InterPro; IPR007114; MFS.
DR InterPro; IPR005828; Sub_transporter.
DR InterPro; IPR001411; Tcr_TetB.
DR Pfam; PF00083; sugar tr; 1.
DR PRINTS; PR01036; TCRTETB.
DR PROSITE; PS0850; MFS; 1.
DR Hypothetical protein; Transport; Transmembrane; Inner membrane;
KW Complete proteome.
FT TRANSMEM 13 33 POTENTIAL.
FT TRANSMEM 49 69 POTENTIAL.
FT TRANSMEM 77 97 POTENTIAL.
FT TRANSMEM 138 158 POTENTIAL.
FT TRANSMEM 165 185 POTENTIAL.
FT TRANSMEM 197 217 POTENTIAL.
FT TRANSMEM 225 245 POTENTIAL.
FT TRANSMEM 263 283 POTENTIAL.
FT TRANSMEM 286 306 POTENTIAL.
FT TRANSMEM 342 362 POTENTIAL.
FT TRANSMEM 396 416 POTENTIAL.
FT TRANSMEM 431 451 POTENTIAL.
SQ SEQUENCE 471 AA; 50901 MW; FCAF02AD330B5F42 CRC64;

Query Match 41.1%; Score 41.5; DB 1; Length 471;
Best Local Similarity 55.6%; Pred. No. 44;
Matches 10; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

QY 3 RAVRLALSG-FDGRVGS 19
DB 261 RTPSLGLAGSFAGRIGSG 278

RESULT 18
YMP3_STRCO
ID YMP3_STRCO STANDARD; PRT; 251 AA.
AC P43168;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Hypothetical oxidoreductase in mprA 5' region (EC 1.-.-.-) (ORF3).
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=DSM 3030 / Mueller;
RX MEDLINE=93023855; PubMed=1406267;
RA Dammann T., Wohlleben W.;
RT "A metalloprotease gene from Streptomyces coelicolor 'Muller' and its
transcriptional activator, a member of the LysR family.";
RL Mol. Microbiol. 6:2267-2278(1992).
CC -!- PATHWAY: May be involved in the biosynthesis of a heptaene-type
antibiotic.
CC -!- SIMILARITY: Belongs to the short-chain dehydrogenases/reductases
(SDR) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z11929; CAA77984.1; -.
DR PIR; S25186; S25186.
DR InterPro; IPR002198; ADH_short.
DR Pfam; PF00106; adh_short; 1.
DR PROSITE; PS00061; ADH_SHORT; 1.
KW Hypothetical protein; Antibiotic biosynthesis; Oxidoreductase.
FT NP_BIND 14 37 NAD OR NADP (BY SIMILARITY).
FT ACT_SITE 158 158 BY SIMILARITY.
SQ SEQUENCE 251 AA; 25777 MW; AF39BF3E53B58A01 CRC64;

Query Match 40.6%; Score 41; DB 1; Length 251;
Best Local Similarity 50.0%; Pred. No. 2;
Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 4 AVRLLALSGFDGRVGS 19
DB 110 AVRVAQTNYTGAGAG 125

RESULT 19
DAPB_STRMU
ID DAPB_STRMU STANDARD; PRT; 255 AA.
AC Q8DUL9;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Dihydrodipicolinate reductase (EC 1.3.1.26) (DHPR).
GN DAPB OR SMU.900.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UAI59 / ATCC 700610 / Serotype C;
RX MEDLINE=22295063; PubMed=12397186;
RA Ajdic D., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y.,
RA Li S., Zhu H., Najjar F., Lai H., White J., Roe B.A., Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental
pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
CC -!- CATALYTIC ACTIVITY: 2,3,4,5-tetrahydrodipicolinate + NAD(P) (+) =
2,3-dihydrodipicolinate + NAD(P)H.
CC -!- PATHWAY: Biosynthesis of diaminopimelate and lysine from aspartate
semialdehyde; second step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the dihydrodipicolinate reductase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC

```



```

RESULT 22
G3P_PSEAE STANDARD; PRT; 334 AA.
AC P27726;
DT 01-AUG-1992 (Rel. 23, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glyceraldehyde 3-phosphate dehydrogenase (EC 1.2.1.12) (GAPDH).
GN GAP OR HEXC OR PA3195.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP 'SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=94321343; PubMed=8045900;
RA Temple L.M., Sage A., Christie G.E., Phibbs P.V. Jr.;
RT "Two genes for carbohydrate catabolism are divergently transcribed
RT from a region of DNA containing the hexC locus in Pseudomonas
RT aeruginosa PA01."
RL J. Bacteriol. 176:4700-4709(1994).
RN [2]
RP 'SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
CC -!- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate + phosphate +
CC NAD(+) = 3-phospho-D-glyceroyl phosphate + NADH.
CC -!- PATHWAY: Second phase of glycolysis; first step.
CC -!- SUBUNIT: Homotetramer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the glyceraldehyde 3-phosphate
CC dehydrogenase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M74256; AAA03488.1; -.
CC DR EMBL; AS004743; AAG0583.1; -.
CC DR PIR; H83246; H83246.
CC DR HSSP; P17721; 1HDG.
CC DR InterPro; IPR000173; GAP dhdrogenase.
CC DR InterPro; IPR006424; GAPDH-I.
CC DR Pfam; PF00044; gpdh; 1.
CC DR Pfam; PF02800; gpdh; 1.
CC DR PRINTS; PR00078; G3PDHDRGNASE.
CC DR TIGRFAM; TIGR01534; GAPDH-I; 1.
CC DR PROSITE; PS00071; GAPDH; 1.
CC DR GlycoLysis; Oxidoreductase; NAD; Complete proteome.
KW BINDING 154 154 GLYCERALDEHYDE 3-PHOSPHATE.
FT ACT_SITE 181 181 ACTIVATES THIOL GROUP DURING CATALYSIS.
FT CONFLICT 140 142 EVL -> RGWV (IN REF. 1).
FT CONFLICT 268 268 E -> K (IN REF. 1).
SQ SEQUENCE 334 AA; 36170 MW; 57A5E2D257DAFF26 CRC64;

Query Match 40.1%; Score 40.5; DB 1; Length 334;
Best Local Similarity 61.5%; Pred. No. 44;
Matches 8; Conservative 4; Mismatches 0; Indels 1; Gaps 1;

```

```

QY 5 VRLALSGFDGRVG 17
DB :||:|||||:|
3 IRLAINGP-GRIG 14

RESULT 23
G3PA_CHLRE STANDARD; PRT; 374 AA.
AC P50362;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Glyceraldehyde 3-phosphate dehydrogenase A, chloroplast precursor
DE (EC 1.2.1.13) (NADP-dependent glyceraldehydephosphate dehydrogenase
DE subunit A).
GN GAP.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP 'SEQUENCE FROM N.A.
RC STRAIN=1132;
RX MEDLINE=94159107; PubMed=8114942;
RA Kersanach R., Brinkmann H., Liaud M.-F., Zhang D.X., Martin W.,
RA Cerff R.;
RT "Five identical intron positions in ancient duplicated genes of
RT subacterial origin."
RL Nature 367:387-389(1994).
RN [2]
RP 3D-STRUCTURE MODELING OF 35-374.
RA Li A.D., Stevens F.J., Huppe H.C., Kersanach R., Anderson L.E.;
RT "Chlamydomonas reinhardtii NADP-linked glyceraldehyde-3-phosphate
RT dehydrogenase contains the cysteine residues identified as potentially
RT domain-locking in the higher plant enzyme and is light activated."
RL Photosyn. Res. 53:167-177(1997).
CC -!- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate + phosphate +
CC NADP(+) = 3-phospho-D-glyceroyl phosphate + NADPH.
CC -!- PATHWAY: Calvin cycle.
CC -!- SUBUNIT: Tetramer of either four A chains (GAPDH 2) or two A and
CC two B chains (GAPDH 1) (By similarity).
CC -!- SUBCELLULAR LOCATION: Chloroplast.
CC -!- MISCELLANEOUS: Plants contain three forms of GAPDH: a cytosolic
CC form which participates in glycolysis and two chloroplast forms
CC which participate in photosynthesis. These three forms are
CC encoded by distinct genes.
CC -!- SIMILARITY: Belongs to the glyceraldehyde 3-phosphate
CC dehydrogenase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L27668; AAA86855.1; -.
CC DR PIR; T07990; T07990.
CC DR PDB; 1NLH; 14-OCT-96.
CC DR PDB; 1NLH; 14-OCT-96.
CC DR InterPro; IPR000173; GAP dhdrogenase.
CC DR InterPro; IPR006424; GAPDH-I.
CC DR Pfam; PF00044; gpdh; 1.
CC DR Pfam; PF02800; gpdh; 1.
CC DR PRINTS; PR00078; G3PDHDRGNASE.
CC DR TIGRFAM; TIGR01534; GAPDH-I; 1.
CC DR PROSITE; PS00071; GAPDH; 1.
CC DR Calvin cycle; Oxidoreductase; NADP; Transit peptide; Chloroplast;
KW Multigene family; 3D-structure.
FT TRANSIT 1 34 CHLOROPLAST (BY SIMILARITY).
FT CHAIN 35 374 GLYCERALDEHYDE 3-PHOSPHATE DEHYDROGENASE
FT A.

```

```

FT BINDING 190 190 GLYCERALDEHYDE 3-PHOSPHATE.
FT ACT_SITE 217 217 ACTIVATES THIOL GROUP DURING CATALYSIS.
FT DISULFID 55 325
FT STRAND 39 43
FT TURN 45 47
FT HELIX 48 54
FT TURN 55 60
FT TURN 65 71
FT HELIX 75 83
FT STRAND 84 85
FT TURN 86 88
FT STRAND 89 90
FT STRAND 103 104
FT STRAND 109 110
FT STRAND 113 114
FT TURN 118 119
FT HELIX 123 126
FT TURN 127 127
FT STRAND 130 133
FT HELIX 141 144
FT TURN 145 145
FT HELIX 146 151
FT STRAND 154 157
FT TURN 171 173
FT TURN 180 181
FT STRAND 184 186
FT HELIX 190 205
FT TURN 206 206
FT STRAND 208 217
FT HELIX 233 236
FT TURN 239 241
FT STRAND 244 246
FT HELIX 250 257
FT TURN 259 260
FT STRAND 265 271
FT STRAND 278 290
FT TURN 292 293
FT HELIX 294 302
FT TURN 303 310
FT STRAND 311 314
FT HELIX 321 323
FT TURN 324 325
FT STRAND 330 333
FT TURN 334 336
FT STRAND 338 339
FT STRAND 342 351
FT TURN 353 354
FT HELIX 355 371
SQ SEQUENCE 374 AA; 40304 MW; 8CEB02897930D34C CRC64;

Query Match 40.1%; Score 40.5; DB 1; Length 374;
Best Local Similarity 43.8%; Pred. No. 50;
Matches 7; Conservative 7; Mismatches 1; Indels 1; Gaps 1;

QY 2 ORAVRLALSGDGRVG 17
Db :::|:|:|:|:|:|:|:|
35 EKIRVAINGF-GRIG 49

RESULT 24
CARR_BUCAI STANDARD; PRT; 1078 AA.
AC P57244;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
phosphate synthetase ammonia chain).
GN CARB OR BU144.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrtosiphon pisum
symbiotic bacterium).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Buchnera.

NCBI_TaxID=118099;
[1]
SEQUENCE FROM N.A.
STRAIN=Tokyo 1998; PubMed=10993077;
MEDLINE=20445173; Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
Shigenobu S., "Genome sequence of the endocellular bacterial symbiont of aphids
Buchnera sp. APS.";
Nature 407:81-86(2000).
-|- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
phosphate + L-glutamate + carbamoyl phosphate.
-|- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
-|- PATHWAY: Arginine biosynthesis.
-|- SUBUNIT: Pyrimidine biosynthesis; first step.
-|- SUBUNIT: Composed of two chains; the small (or glutamine) chain
promotes the hydrolysis of glutamine to ammonia, which is used by
the large (or ammonia) chain to synthesize carbamoyl phosphate (By
similarity).
-|- SIMILARITY: Belongs to the carB family.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AP001119; BAB12862.1; -.
DR HSSP; P00968; 1CS0.
DR HAMAP; MF_01210; -. 1.
DR InterPro; IPR006275; CarA_L_glu.
DR InterPro; IPR005483; CPase_L_D2.
DR InterPro; IPR005479; CPase_L_D3.
DR InterPro; IPR005480; CPase_L_D3.
DR InterPro; IPR005481; CPase_L_N.
DR InterPro; IPR004362; MGS like.
DR Pfam; PF00289; CPase_L_Chain; 2.
DR Pfam; PF02786; CPase_L_D2; 2.
DR Pfam; PF02787; CPase_L_D3; 1.
DR Pfam; PF02142; MGS; 1.
DR PRINTS; PR0098; CPASE.
DR TIGRFAMs; TIGR01369; CPaseII_lrg; 1.
DR PROSITE; PS00866; CPASE_1; 2.
DR PROSITE; PS00867; CPASE_2; 2.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
ATP-binding; Manganese; Complete proteome.
INIT MET 0
FT DOMAIN 1 402 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 403 552 OLIGOMERIZATION DOMAIN.
FT DOMAIN 553 935 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 936 1078 ALLOSTERIC DOMAIN.
FT REPEAT 1 552
FT REPEAT 553 1078
FT NP_BIND 152 209 ATP (POTENTIAL).
FT NP_BIND 302 353 ATP (POTENTIAL).
FT METAL 284 284 MANGANESE 1 (BY SIMILARITY).
FT METAL 298 298 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 300 300 MANGANESE 2 (BY SIMILARITY).
FT METAL 828 828 MANGANESE 3 (BY SIMILARITY).
FT METAL 840 840 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1078 AA; 120384 MW; 2F5060604D540901 CRC64;

Query Match 40.1%; Score 40.5; DB 1; Length 1078;
Best Local Similarity 38.1%; Pred. No. 1.5e+02;
Matches 8; Conservative 6; Mismatches 4; Indels 3; Gaps 1;

QY 1 LQRAVR---LALSGDGRVGS 18
Db :|:|:|:|:|:|:|:|
394 IQAIRGLEVGASGDFSKISS 414

RESULT 25

```

MODB ECOLI
 ID MODB_ECOLI STANDARD; PRT; 229 AA.
 AC P09834; P77537;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Molybdenum transport system permease protein modB.
 GN MODB OR CHLJ OR B0764 OR Z0934 OR ECS0792.
 OS Escherichia coli, and
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562, 83334;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RC MEDLINE=95394784; PubMed=7665460;
 RA Maupin-Furlow J.A., Rosentel J.K., Lee J.H., Deppenmeier U.,
 RA Gunsalus R.P., Shammugam K.T.;
 RT "Genetic analysis of the modABCD (molybdate transport) operon of
 RT Escherichia coli.";
 RL J. Bacteriol. 177:4851-4856(1995).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=K12 / MC1000;
 RC MEDLINE=96151473; PubMed=8564363;
 RA Walkenhorst H.M., Hemschemeyer S.K., Eichenlaub R.;
 RT "Molecular analysis of the molybdate uptake operon, modABCD, of
 RT Escherichia coli and modR, a regulatory gene.";
 RL Microbiol. Res. 150:347-361(1995).
 RN [3]
 RN SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RC MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [4]
 RN SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RC MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayaashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Maeda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
 RA Yano M., Horiuchi T.;
 RT "A 718-Kb DNA sequence of the Escherichia coli K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map.";
 RL DNA Res. 3:137-155(1996).
 RN [5]
 RN SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RC MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouisis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
 RL Nature 409:529-533(2001).
 RN [6]
 RN SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RC MEDLINE=21156231; PubMed=11258796;
 RA Hayaashi T., Makino K., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;

RT
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RN DNA Res. 8:11-22(2001).
 RN [7]
 RN SEQUENCE OF 30-229 FROM N.A.
 RC STRAIN=K12;
 RC MEDLINE=87194564; PubMed=3553151;
 RA Johann S., Hinton S.W.;
 RT "Cloning and nucleotide sequence of the chld locus.";
 RL J. Bacteriol. 169:1911-1916(1987).
 CC -!- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM
 CC FOR MOLYBDENUM; PROBABLY RESPONSIBLE FOR THE TRANSLLOCATION OF THE
 CC SUBSTRATE ACROSS THE MEMBRANE.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (potential).
 CC -!- SIMILARITY: Belongs to the binding-protein-dependent transport
 CC system permease family. CysTW subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L34009; AAB00836.1; -
 CC EMBL; U27192; AAB60172.1; -
 CC EMBL; U07867; AAB06894.1; -
 CC EMBL; AE000179; AAC73851.1; -
 CC EMBL; D90715; BAA35428.1; -
 CC EMBL; M16182; AAB83839.1; -
 CC EMBL; AR005254; AAG55093.1; -
 CC EMBL; AP002553; BAB34215.1; -
 CC PIR; A85579; A85579.
 CC PIR; D64812; BVECHJ.
 CC PIR; H90727; H90727.
 CC Ecogene; EG10002; modB.
 CC InterPro; IPR000515; BPD transp.
 CC InterPro; IPR006229; Mol_porter.
 CC Pfam; PF00528; BPD transp; 1.
 CC TIGRFAMs; TIGR01095; modB; 1.
 CC PROSITE; PS50928; ABC_TM1; 1.
 CC Inner membrane; Transmembrane; Molybdenum; Transport;
 KW Complete proteome.
 FT TRANSMEM 17 37 POTENTIAL.
 FT TRANSMEM 50 70 POTENTIAL.
 FT TRANSMEM 84 104 POTENTIAL.
 FT TRANSMEM 137 157 POTENTIAL.
 FT TRANSMEM 202 222 POTENTIAL.
 FT CONFLICT 40 40 T -> S (IN REF. 1, 2 AND 7).
 SQ SEQUENCE 229 AA; 24938 MW; 65A3A05FBF3382C6 CRC64;
 Query Match 39.6%; Score 40; DB 1; Length 229;
 Best Local Similarity 72.7%; Pred. No. 36;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 3 RAVRLALSGFD 13
 Db 109 RAIRLALGVD 119
 ||:|||||
 RESULT 26
 ID DAPB CAUCR STANDARD; PRT; 257 AA.
 AC Q9AZII;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Dihydrodipicolinate reductase (EC 1.3.1.26) (DHPR).
 GN DAPB OR CC3550.
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;

```

OC Caulobacteraceae; Caulobacter.
OX NCBI_TaxID=155892;
RN [1]_SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J.A., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Pocko A.I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoa R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Clavel M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RA "Complete genome sequence of Caulobacter crescentus.";
RT Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
RL CC -1- CATALYTIC ACTIVITY: 2,3,4,5-tetrahydrodipicolinate + NAD(P)(+) =
CC 2,3-dihydrodipicolinate + NAD(P)H.
CC -1- PATHWAY: Biosynthesis of diaminopimelate and lysine from aspartate
CC senialdehyde; second step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the dihydrodipicolinate reductase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE006013; AK25512.1; -.
DR PIR; D87689; D87689.
DR HSSP; P04036; 1DRW.
DR TIGR; CC3550; -.
DR HAMAP; MF_00102; -.
DR InterPro; IPR000846; DapB.
DR Pfam; PF05173; DapB_C; 1.
DR Pfam; PF01113; DapB_N; 1.
DR ProDom; PD004105; DapB; 1.
DR TIGRFams; TIGR00036; dapB; 1.
DR PROSITE; PS01298; DAPB; 1.
KW Diaminopimelate biosynthesis; Lysine biosynthesis; Oxidoreductase;
KW NADP; Complete proteome.
SQ SEQUENCE 257 AA; 26425 MW; B8AD3BB269A47AE4 CRC64;

Query Match 39.6%; Score 40; DB 1; Length 257;
Best Local Similarity 35.3%; Pred. No. 41; Mismatches 3; Indels 0; Gaps 0;
Matches 6; Conservative 8;

QY 1 LQRAVRLALSGFDGRVG 17
DB :|:|:|:|:|:|
1 MSQPVKIATAGNRWG 17

RESULT 27
YQ46 STRCO STANDARD; PRT; 279 AA.
AC Q9Z596;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein SC06206.
GN SC06206 OR SC205-27C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]_SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,

```

```

RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA Rabinowitch E., Rajandream M.A., Rutherford K., Rutter S., Taylor K.,
RA Seeger K., Saunders D., Sharp S., Squares S., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RA "Complete genome sequence of the model actinomycete Streptomyces
RA coelicolor A3(2).";
RT Nature 417:141-147(2002).
RL CC -1- SIMILARITY: BELONGS TO THE HYI FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL939126; CAB36614.1; -.
DR PIR; T34860; T34860.
DR InterPro; IPR001719; AP endonuclease2.
DR Pfam; PF01261; AP endonuc 2; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 279 AA; 29635 MW; 18010C6C3A755D73 CRC64;

Query Match 39.6%; Score 40; DB 1; Length 279;
Best Local Similarity 66.7%; Pred. No. 44;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 6 RLALSGFDGRVG 17
DB :|:|:|:|:|:|
243 RLAKAGYDGVWG 254

RESULT 28
MOAA AERPE STANDARD; PRT; 355 AA.
AC Q9YEV3;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable molybdenum cofactor biosynthesis protein A.
GN MOAA OR APE0478.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococccaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]_SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RA "Complete genome sequence of an aerobic hyper-thermophilic
RA crenarchaeon, Aeropyrum pernix K1.";
RT DNA Res. 6:83-101(1999).
RL CC -1- FUNCTION: Together with moaC, is involved in the conversion of a
CC guanosine derivative (GMP) into molybdopterin precursor Z (By
CC similarity).
CC -1- COFACTOR: Binds 1 Fe-4S cluster (By similarity).
CC -1- PATHWAY: Molybdenum cofactor biosynthesis; first step.
CC -1- SIMILARITY: Belongs to the radical SAM superfamily. MoaA family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

modified and this statement is not removed. Usage by and for commercial
or send an email to license@isb-sib.ch) (See <http://www.isb-sib.ch/announce/>)

CC EMBL; AP000059; BAA79443.1; -
CC PIR; G72743; G72743. -
CC HAMAP; MF_01225; -; 1.
CC InterPro; IPR006638; Elp3.
CC InterPro; IPR000385; MoaA_NifB_PqqE.
CC InterPro; IPR007197; Radical SAM.
CC Pfam; PF04055; Radical SAM; 1.
CC SMART; SM00729; Elp3; 1.
CC PROSITE; PS01305; MOA_NIFB_PQQE; 1.
CC Molybdenum cofactor biosynthesis; Metal-binding; Iron; Iron-sulfur;
KW 3Fe-4S; Complete proteome.
FT METAL 21 IRON-SULFUR (3FE-4S) (POTENTIAL).
FT METAL 28 IRON-SULFUR (3FE-4S) (POTENTIAL).
FT METAL 252 IRON-SULFUR (3FE-4S) (POTENTIAL).
FT METAL 269 IRON-SULFUR (3FE-4S) (POTENTIAL).
SQ SEQUENCE 355 AA; 39488 MW; 5570394978631788 CRC64;

Query Match 39.6%; Score 40; DB 1; Length 355;
Best Local Similarity 44.4%; Pred. No. 57;
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 2 QRAVRLALSGFDGRVSG 19
DB 136 RRAVKLALKGVEAALRAG 153
:|||||: : : : :

RESULT 29
DCOR DATST STANDARD; PRT; 431 AA.
AC P50134;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Ornithine decarboxylase (EC 4.1.1.17) (ODC).
OS Datura stramonium (Jimsonweed) (Common thornapple).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC lamids; Solanales; Solanaceae; Datura.
OX NCBI_TaxID=4076;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. D15/5; TISSUE=Root;
RX MEDLINE=96195135; PubMed=8660289;
RA Michael A.J., Furze J.M., Rhodes M.J.C., Burtin D.;
RT "Molecular cloning and functional identification of a plant ornithine
decarboxylase cDNA.";
RL Biochem. J. 314:241-248(1996).
CC -!- CATALYTIC ACTIVITY: L-ornithine = putrescine + CO(2).
CC -!- COFACTOR: Pyridoxal phosphate.
CC -!- PATHWAY: Polyamine biosynthesis; first (rate-limiting) step.
CC -!- SIMILARITY: BELONGS TO FAMILY 2 OF ORNITHINE, DAP, AND ARGinine
DECARBOXYLASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
or send an email to license@isb-sib.ch)

CC EMBL; X87847; CAA61121.1; -
CC PIR; S64704; S64704.
CC HSSP; P07805; 1F3T.
CC InterPro; IPR000183; Decarboxylase2.
CC InterPro; IPR009006; Racem decarbox_C.
CC Pfam; PF02784; Orn_Arg_deC_N; 1.
CC Pfam; PF00278; Orn_DAP_Arg_deC; 1.
CC PRINTS; PR01179; ODACRBXLASE.

DR PROSITE; PS00878; ODR_DC_2_1; 1.
DR PROSITE; PS00879; ODR_DC_2_2; 1.
KW Lyase; Decarboxylase; Pyridoxal phosphate; Polyamine biosynthesis;
KW Phosphorylation.
FT BINDING 94 94 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT ACT_SITE 376 376 BY SIMILARITY.
SQ SEQUENCE 431 AA; 46659 MW; 8168064D8CC1C9F8 CRC64;

Query Match 39.6%; Score 40; DB 1; Length 431;
Best Local Similarity 47.1%; Pred. No. 69;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 3 RAVRLALSGFDGRVSG 19
DB 211 QAARLTGVGVSHIGSG 227
:|||||: : : : :

RESULT 30
SPS2_HUMAN STANDARD; PRT; 448 AA.
AC Q99611; Q9BUQ2;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Selenide, water dikinase 2 (EC 2.7.9.3) (Selenophosphate synthetase 2)
DE (Selenium donor protein 2).
GN SEPHS2 OR SPS2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96017645; PubMed=7588067;
RA Guimaraes M.J., Bazan J.F., Zlotnik A., Wiles M.V., Grimaldi J.C.,
RA Lee F., McClanahan T.;
RT "A new approach to the study of haematopoietic development in the
yolk sac and embryoid bodies.";
RL Development 121:3335-3346(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97140286; PubMed=8986768;
RA Guimaraes M.J., Peterson D., Vicari A., Cocks B.G., Copeland N.G.,
RA Gilbert D.J., Jenkins N.A., Ferrick D.A., Kastelein R., Bazan J.F.,
RA Zlotnik A.;
RT "Identification of a novel seld homolog from eukaryotes, bacteria,
and archaea: is there an autoregulatory mechanism in selenocysteine
metabolism?";
RL Proc. Natl. Acad. Sci. U.S.A. 93:15086-15091(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle, and Skin;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg K.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner K.H., Schaefer C.P., Bhat N.K.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Hellon E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalek U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

CC -!- FUNCTION: Synthesizes selenophosphate from selenide and ATP.
CC -!- CATALYTIC ACTIVITY: ATP + selenide + H(2)O = AMP + selenophosphate
CC + phosphate.
CC -!- COFACTOR: Selenocysteine. The active-site selenocysteine is
CC encoded by the opal codon, UGA.
CC -!- SIMILARITY: Belongs to the selenophosphate synthetase 1 family.
CC Class I subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U43286; AAC50958.2; -.
DR EMBL; BC002381; AAH02381.3; -.
DR EMBL; BC016643; AAH16643.1; -.
DR MIM; 606218; -.
DR GO; GO:0004756; F:selenide, water dikinase activity; NAS.
DR GO; GO:0016260; P:selenocysteine biosynthesis; NAS.
DR InterPro; IPR000728; AIR synth.
DR InterPro; IPR004536; SelD.
DR Pfam; PF0586; AIRS; 1.
DR Pfam; PF02769; AIRS_C; 1.
DR TIGRFAMs; TIGR00476; selD; 1.
KW Transferase; Selenium; Selenocysteine; ATP-binding.
FT ACT_SITE 60 60
FT SE_CYS 60 60
FT SITE 63 63
FT SITE 63 63
FT SITE 63 63
FT NP_BIND 319 325
FT SEQUENCE 448 AA; 47258 MW; 343A58CD9F842B99 CRC64;
SQ
Query Match 39.6%; Score 40; DB 1; Length 448;
Best Local Similarity 58.3%; Pred. No. 72;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 9 LSGFDGRVSGC 20
Db 51 LTGFGMGKGC 62

Search completed: May 13, 2004, 06:55:48
Job time : 9.66667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2004, 06:49:47 ; Search time 33.3333 Seconds
(without alignments)
189.311 Million cell updates/sec

Title: 09549186-8

Perfect score: 101

Sequence: 1 LQRAVRLALSGDGRVSGC 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database :

SPREMBL_25.*

1: sp_archaea.*

2: sp_bacteria.*

3: sp_fungi.*

4: sp_human.*

5: sp_invertebrate.*

6: sp_mammal.*

7: sp_mhc.*

8: sp_organelle.*

9: sp_phase.*

10: sp_plant.*

11: sp_rodent.*

12: sp_virus.*

13: sp_vertebrate.*

14: sp_unclassified.*

15: sp_rvirus.*

16: sp_bacteriap.*

17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	48	47.5	299	16	Q98KT8
2	48	47.5	438	2	Q845V3
3	47	46.5	359	16	Q910G7
4	47	46.5	575	17	Q28180
5	47	46.5	606	10	Q8LMR1
6	46	45.5	166	10	Q8W5C0
7	46	45.5	373	10	Q7XKP6
8	46	45.5	379	10	Q7XMO0
9	46	45.5	405	10	Q9AYM0
10	46	45.5	405	10	Q7X976
11	45.5	45.0	473	16	Q916M2
12	45.5	45.0	1482	16	Q8EHA4
13	45	44.6	453	10	Q9LKR6
14	45	44.6	484	10	Q8LQF2
15	45	44.6	499	5	Q86B00
16	45	44.6	503	16	Q8UBB2

17	44.5	44.1	408	16	Q7VGM3
18	44	43.6	126	4	Q9UHE2
19	44	43.6	136	16	Q98J09
20	44	43.6	136	16	Q92ZC0
21	44	43.6	226	4	Q9Y5T7
22	44	43.6	485	4	Q96GW2
23	43.5	43.1	325	2	Q8RMH1
24	43.5	43.1	414	10	Q93YH6
25	43.5	43.1	596	16	Q8PHP2
26	43	42.6	147	16	Q82IR7
27	43	42.6	254	16	Q81S30
28	43	42.6	254	16	Q81FG5
29	43	42.6	287	16	Q83611
30	43	42.6	352	16	Q92AC6
31	43	42.6	352	16	Q8Y619
32	43	42.6	383	16	Q82CN1
33	43	42.6	388	16	Q9WH7
34	43	42.6	393	5	Q9V353
35	43	42.6	404	2	Q93AD0
36	43	42.6	477	17	Q9HRL0
37	43	42.6	596	5	Q9V766
38	43	42.6	688	5	Q8T9L6
39	43	42.6	1221	16	Q910U2
40	43	42.6	1227	5	Q9W0M7
41	43	42.6	1997	10	Q8LRM7
42	42.5	42.1	229	16	Q7VNY4
43	42.5	42.1	295	2	Q8KSX8
44	42.5	42.1	331	2	Q9X7H6
45	42.5	42.1	335	16	Q9A3F6
46	42.5	42.1	335	16	Q8Y1X9
47	42.5	42.1	335	16	Q8FYX9
48	42.5	42.1	338	16	Q8ZHH2
49	42.5	42.1	407	5	Q27661
50	42.5	42.1	622	16	Q8D7H0
51	42.5	42.1	1253	16	Q88H58
52	42	41.6	50	13	Q8JFC6
53	42	41.6	51	13	Q8J299
54	42	41.6	54	13	Q8JFE8
55	42	41.6	135	16	Q8UAA7
56	42	41.6	215	16	Q82CX3
57	42	41.6	229	16	Q8FJR5
58	42	41.6	262	16	Q7V8V6
59	42	41.6	271	10	Q851F8
60	42	41.6	280	2	Q9APF1
61	42	41.6	435	16	Q8EKL8
62	42	41.6	455	16	Q7VE01
63	42	41.6	465	12	Q9QDK9
64	42	41.6	466	10	Q9SNC9
65	42	41.6	472	16	Q8D9F3
66	42	41.6	478	16	Q87PA5
67	42	41.6	497	16	Q93JB9
68	42	41.6	514	10	Q7XNU9
69	42	41.6	799	16	Q8G5U6
70	42	41.6	862	5	Q22354
71	42	41.6	940	2	Q51187
72	42	41.6	943	2	Q87343
73	42	41.6	943	2	Q50952
74	42	41.6	1037	5	Q9VVR9
75	42	41.6	1151	5	O44319
76	42	41.6	1399	16	Q826G9
77	42	41.6	1471	2	Q8KKT7
78	42	41.6	1581	16	Q8U9B4
79	42	41.6	1865	11	Q60544
80	41.5	41.1	183	2	Q9LC99
81	41.5	41.1	337	16	Q83HP7
82	41.5	41.1	337	16	Q83G13
83	41.5	41.1	471	16	Q8X713
84	41.5	41.1	471	16	Q8FG02
85	41.5	41.1	471	16	Q83QZ0
86	41.5	41.1	501	16	Q9WYV3
87	41.5	41.1	689	10	Q9T0B2
88	41.5	41.1	747	10	Q22990
89	41	40.6	104	2	Q50013

Q7vgm3 helicobacte
Q9uhe2 homo sapien
Q98j09 rhizobium l
Q92zc0 rhizobium m
Q9y5t7 homo sapien
Q96gw2 homo sapien
Q8rmh1 acetobacter
Q93yh6 galdieria s
Q8php2 xanthomonas
Q83ir7 streptomyce
Q82ir7 streptomyce
Q81s30 bacillus an
Q81fg5 bacillus ce
Q83611 treponema p
Q92ac6 listeria in
Q8y619 listeria mo
Q82cn1 streptomyce
Q9wh7 pseudomonas
Q9v353 drosophila
Q93ad0 pseudomonas
Q9hr10 halobacteri
Q9v766 drosophila
Q8t9l6 drosophila
Q910u2 pseudomonas
Q9w0m7 drosophila
Q8lrn7 chlamydomon
Q7vny4 haemophilus
Q8ksx8 streptomyce
Q9x7h6 paracoccus
Q9ajf6 caulobacter
Q8yix9 bruceella me
Q8fyx9 bruceella su
Q8zhh2 yersinia pe
Q27661 giardia lam
Q8d7h0 vibrio vuln
Q88h58 pseudomonas
Q8jfc6 ficedula ny
Q8j299 ficedula al
Q8jfe8 ficedula al
Q8uaa7 agrobacteri
Q82cx3 streptomyce
Q8fjr5 escherichia
Q7v8v6 prochloroco
Q851f8 oryza sativ
Q9apf1 xanthomonas
Q8ekl8 shewanella
Q7ve01 prochloroco
Q9qdk9 pterostyllis
Q9enc9 arabidopsis
Q8d9f3 vibrio vuln
Q87pa5 vibrio para
Q93jb9 streptomyce
Q7xnu9 oryza sativ
Q8g5u6 bifidobacte
Q22354 caenorhabdi
Q51187 neisseria m
Q87343 neisseria m
Q50952 neisseria g
Q9vvr9 drosophila
O44319 anurida mar
Q826g9 streptomyce
Q8kkt7 rhizobium e
Q8u9e4 agrobacteri
Q60544 mesocricetu
Q9lc99 bacillus ha
Q83hp7 tropheryma
Q8jg13 tropheryma
Q8x713 escherichia
Q8fg02 escherichia
Q83qz0 shigella fl
Q9wyv3 thermotoga
Q9t0b2 arabidopsis
Q22990 arabidopsis
Q50013 mycobacteri

90	41	40.6	108	3	Q8TFP0
91	41	40.6 <td>114</td> <td>1</td> <td>Q8U4U3</td>	114	1	Q8U4U3
92	41	40.6 <td>117</td> <td>17</td> <td>Q8PYW7</td>	117	17	Q8PYW7
93	41	40.6 <td>143</td> <td>16</td> <td>Q8UPP7</td>	143	16	Q8UPP7
94	41	40.6 <td>203</td> <td>16</td> <td>Q7WDU4</td>	203	16	Q7WDU4
95	41	40.6 <td>203</td> <td>16</td> <td>Q7W2L8</td>	203	16	Q7W2L8
96	41	40.6 <td>236</td> <td>16</td> <td>Q92KE3</td>	236	16	Q92KE3
97	41	40.6 <td>240</td> <td>16</td> <td>Q83A70</td>	240	16	Q83A70
98	41	40.6 <td>246</td> <td>16</td> <td>Q89TS9</td>	246	16	Q89TS9
99	41	40.6 <td>255</td> <td>16</td> <td>Q8DUL9</td>	255	16	Q8DUL9
100	41	40.6 <td>256</td> <td>5</td> <td>Q8SUF1</td>	256	5	Q8SUF1
					Q8TFP0 tricholoma
					Q8U4U3 methanosarc
					Q8PYW7 methanosarc
					Q8UPP7 agrobacteri
					Q7WDU4 bordetella
					Q7W2L8 bordetella
					Q92KE3 rhizobium m
					Q83A70 coxiella bu
					Q89TS9 bradyrhizob
					Q8DUL9 streptococc
					Q8SUF1 encephalito

ALIGNMENTS

RESULT 1	Q98KTB	PRELIMINARY;	PRT;	299 AA.
ID	Q98KTB			
AC	Q98KTB			
DT	01-OCT-2001	(T-EMBLrel. 18, Created)		
DT	01-OCT-2001	(T-EMBLrel. 18, Last sequence update)		
DT	01-OCT-2003	(T-EMBLrel. 25, Last annotation update)		
DE	Succinyl-CoA synthetase alpha subunit.			
GN	MLR1326.			
OS	Rhizobium loti (Mesorhizobium loti).			
OC	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;			
OC	Phyllobacteriaceae; Mesorhizobium.			
OX	NCBI_TaxId=381;			
OX	[1]			
RP	SEQUENCE FROM N.A.			
RP	STRAIN=MAFF303099;			
RC	MEDLINE=21082930; PubMed=11214968;			
RC	Kaneko T., Nakamura Y., Sato S., Aeamizu E., Kato T., Sasamoto S.,			
RA	Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,			
RA	Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,			
RA	Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,			
RA	Takeuchi C., Yamada M., Tabata S.;			
RT	"Complete genome structure of the nitrogen-fixing symbiotic bacterium			
RT	Mesorhizobium loti.";			
RL	DNA Res. 7:331-338(2000).			
RL	EMBL; AP002997; BAB48726.1; -.			
DR	GO; GO:0003824; P:catalytic activity; IEA.			
DR	GO; GO:0008152; P:metabolism; IEA.			
DR	InterPro; IPR003781; CoA binding.			
DR	InterPro; IPR005811; CoA_ligase.			
DR	InterPro; IPR005810; CoA_lig_alpha.			
DR	Pfam; PF02629; CoA binding; 1.			
DR	Pfam; PF00549; ligase-CoA; 1.			
DR	PRINTS; PR01798; SCOSYNTHASE.			
DR	TIGRFAMs; TIGR01019; succoalalpha; 1.			
DR	PROSITE; PS00399; SUCCINYL_COA_LIG_2; 1.			
DR	Complete proteome.			
SK	SEQUENCE 299 AA; 31601 MW; 78194A84A92079FF CRC64;			

OS	Burkholderia multivorans.
OC	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC	Burkholderiaceae; Burkholderia.
NCBI_TaxID=87883;	
[1]	
RN	SEQUENCE FROM N.A.
RP	STRAIN=ATCC 17616;
RC	Komatsu H., Imura Y., Ohori A., Nagata Y., Tsuda M.;
RA	"Distribution and Organization of Auxotrophic Genes on Multi-
RT	Chromosomal Genome of Burkholderia multivorans ATCC17616.";
RL	Submitted (SEP-2002) to the EMBL/GenBank/DDJB databases.
DR	EMBL; AB091436; BAC65270.1; -.
DR	GO; GO:0004399; F:histidinol dehydrogenase activity; IEA.
DR	GO; GO:0000305; P:histidine biosynthesis; IEA.
DR	InterPro; IPR001692; Histidinol dh.
DR	Pfam; PF00815; Histidinol dh; 1.
DR	PRINTS; PR00083; HOLDHRRGNASE.
DR	PRODOM; PD002680; Histidinol_dh; 1.
DR	TIGRFAMs; TIGR00069; hisd; 1.
SQ	SEQUENCE 438 AA; 46606 MW; 6D383BE84CDB5E25 CRC64;
Query Match	47.5%; Score 48; DB 2; Length 438;
Best Local Similarity	47.8%; Pred. No. 34;
Matches 11; Conservative 3; Mismatches 5; Indels 4;	
QY	2 QAVRLALSGFD----GRVGSGC 20
DB	74 QDALQAALDGLEPKARGAGSGC 96
RESULT 3	
Q9IOG7	PRELIMINARY; PRT; 359 AA.
ID	Q9IOG7
AC	DQ10G7;
DT	01-MAR-2001 (TrEMBLrel. 16, Created)
DI	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE	Hypothetical protein PA2671.
GN	PA2671.
OS	Pseudomonas aeruginosa.
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales
OC	Pseudomonadaceae; Pseudomonas.
NCBI_TaxID=287;	
[1]	
RN	SEQUENCE FROM N.A.
RP	STRAIN=ATCC 15692 / PA01;
RC	MEDLINE=20437337; PubMed=10984043;
RX	Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren
RA	Hickey M.J., Brinkman F.S.L., Huffnagle W.O., Kowalik D.J., Lag
RA	Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yua
RA	Brody L.L., Coulter S.N., Folger K.R., Wu Z., Paulsen I.T., Lim
RA	Smith K.A., Spencer D.H., Wong G.K.-S., Ku A., Paulsen I.T.,
RA	Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.; "
RT	"Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT	opportunistic pathogen.";
RL	Nature 406:959-964(2000).
DR	EMBL; AE004695; AAG06059.1; -.
DR	PIR; E83312; E83312.
DR	GO; GO:0005351; F:sugar porter activity; IEA.
DR	GO; GO:0009401; P:phenolpyruvate-dependent sugar phospho
DR	InterPro; IPR002114; HPr_SerP_S.
DR	PROSITE; PS00589; PTS_HPr_SER; 1.
DR	Hypothetical protein; Complete proteome.
SQ	SEQUENCE 359 AA; 39404 MW; ACAD05FA9187B566 CRC64;

RESULT 2					
ID	Q845V3	PRELIMINARY;	PRT;	438 AA.	
AC	Q845V3				
DT	01-JUN-2003	(TREMBLrel. 24, Created)			
DT	01-JUN-2003	(TREMBLrel. 24, Last sequence update)			
DT	01-OCT-2003	(TREMBLrel. 25, Last annotation update)			
DE	Histidinol dehydrogenase.				
GN	HISD.				
GN	HISD.				
<hr/>					
Query Match	46.5%	Score	47;	DB	16;
Best Local Similarity	62.5%;	Pred. No.	39;		
Matches	10;	Conservative	2;	Mismatches	4;
Indels					
Gaps					
OY	1	LQRAYRLALSGFDGRV	16		
Db	126	LDRSARLSLGIDPRV	141		


```

RESULT 4
O28180
ID O28180 PRELIMINARY; PRT; 575 AA.
AC O28180;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Acetolactate synthase, large subunit (ILVB-4).
GN AF2100.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Ariach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370 (1997).
CC -!- SIMILARITY: WITH OTHER ENZYMES WHICH REQUIRE TPP.
DR EMBL; AE000958; AAB89146.1; --
DR PIR; D69512; D69512.
DR HSSP; P07342; LJSC.
DR TIGR; AF2100; --
DR InterPro; IPR000399; Pyruvate_decarb.
DR Pfam; PF00205; TPP_enzymes; 1.
DR Pfam; PF02775; TPP_enzymes_C; 1.
DR Pfam; PF02776; TPP_enzymes_N; 1.
DR PROSITE; PS00187; TPP_ENZYMES; 1.
KW Flavoprotein; Hypothetical protein; Thiamine pyrophosphate;
KW Complete proteome.
SQ SEQUENCE 575 AA; 63000 MW; B121765C66EC388A CRC64;

Query Match 46.5%; Score 47; DB 17; Length 575;
Best Local Similarity 62.5%; Pred. No. 65;
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 LORAVRLALSGFDGRV 16
Db 147 VQRAIRIALSGRGPV 162

RESULT 5
O8LMR1
ID O8LMR1 PRELIMINARY; PRT; 606 AA.
AC O8LMR1;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative serine/threonine protein kinase.
GN QJ1705B08.12.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA Wing R.A., Yu Y., Soderlund C., Chen M., Kim H.-R., Rambo T.,
RA Saski C., Henry D., Oates R., Simmons J.;

"Rice Genomic Sequence.";
Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC105363; RAM52323.1; --
DR Gramene; O8LMR1; --
DR GO; GO:0005524; P-ATP binding; IEA.
DR GO; GO:0004674; P-protein serine/threonine kinase activity; IEA.
DR GO; GO:0004713; P-protein-tyrosine kinase activity; IEA.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro; IPR000719; Prot kinase.
DR InterPro; IPR002290; Ser_thr_kinase.
DR InterPro; IPR008271; Ser_thr_kinase.
DR InterPro; IPR001245; Tyr_kinase; 1.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Prot kinase; 1.
DR SMART; SM00220; S_TKc_1.
DR SMART; SM00219; TyRKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 606 AA; 67638 MW; 1910629AF80E53ED CRC64;

Query Match 46.5%; Score 47; DB 10; Length 606;
Best Local Similarity 64.3%; Pred. No. 69;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 7 LALSGFDGRVSGSC 20
Db 583 LALHGFSGKVGWEC 596

RESULT 6
O8W5C0
ID O8W5C0 PRELIMINARY; PRT; 166 AA.
AC O8W5C0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Putative GTPase regulator protein.
GN OSUNBB0013K08.11.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Buell C.R., Yuan Q., Ouyang S., Liu J., Moffat K.S., Hill J.N.,
RA Ganaberg K., Brenner M., Burgess S., Hance M., Shvartsbeyn M.,
RA Tsitrin T., Riggs P., Heiao J., Zismann V., Blunt S., Pai G.,
RA VanAken S.E., Utterback T.R., Feldlyum T.V., Kaib E., Quackenbush J.,
RA Salzberg S.L., White O., Fraser C.M.;
RT "Oryza sativa chromosome 3 BAC OSUNBB0013K08 genomic sequence.";
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC092390; AAL31680.1; --
DR Gramene; O8W5C0; --
SQ SEQUENCE 166 AA; 18447 MW; 76A28ED4EDEE33C3 CRC64;

Query Match 45.5%; Score 46; DB 10; Length 166;
Best Local Similarity 50.0%; Pred. No. 25;
Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 2 QRAVRLALSGFDGRVSGS 19
Db 89 EKAAARAARKGFDGSGEG 106

RESULT 7
Q7XXP6
ID Q7XXP6 PRELIMINARY; PRT; 373 AA.

```

```

AC Q7XXP6;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Callus;
RA Moriguchi K., Ito Y., Yamazaki Y., Kurata N.;
RT "Finding of various plant nuclear proteins using yeast nuclear
RT transport trap system - a proteomal approach.";
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB110206; BAC78598.1; -.
KW Hypothetical protein.
SQ SEQUENCE 373 AA; 37513 MW; 9B6222F770CFC6DA7 CRC64;

Query Match 45.5%; Score 46; DB 10; Length 373;
Best Local Similarity 57.1%; Pred. No. 59;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 7 LALSGDGRVGSQC 20
:|:|:|:|:|:|
Db 254 VALAGSDGRVGGC 267

RESULT 8
Q7XM00 PRELIMINARY; PRT; 379 AA.
AC Q7XM00;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE OSJNBa0086006.9 protein.
GN OSJNBa0086006.9.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RA Han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X.,
RA Liu Y.L., Mu J., Yu Z., Chen L., Pan D.L., Weng Q.J., Zhang L.,
RA Lu Y.Q., Yu S.L., Liu X.H., Lu T.T., Zhang Y.J., Lu Y., Li C., Li T.,
RA Zhang Y., Hu H., Jia P.X., Qian Y.M., Ying K., Zhou B., Chen Z.H.,
RA Hao P., Zhang L., Wu M., Zhang R.Q., Guan J.P., Fu G., Wang S.Y.,
RA Ren S.X., Lv G., Lin W., Gu W.Q., Zhu G.F., Tu Y.F., Jia J., Yin H.F.,
RA Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y., Shao C.Y., Sun Y.,
RA Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W., Sheng H.H.,
RA Gu J.L., Chen S.T., Ni L., Zhu F.H., Hong G.F.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL662981; CAE04861.1; -.
SQ SEQUENCE 379 AA; 38241 MW; FC493F8D769E4670 CRC64;

Query Match 45.5%; Score 46; DB 10; Length 379;
Best Local Similarity 57.1%; Pred. No. 60;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 7 LALSGDGRVGSQC 20
:|:|:|:|:|:|
Db 254 VALAGSDGRVGGC 267

RESULT 9
Q9AYM0 PRELIMINARY; PRT; 405 AA.
AC Q9AYM0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)

```

```

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative AT-hook DNA-binding protein.
GN OSJNBa0003019.1.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Buell C.R., Yuan Q., Moffat K.S., Hill J.N., Burr P.C., Heiao J.,
RA Zismann V., Pai G., Bowman C.L., Fujii C.Y., VanAken S.E.,
RA Bowman C.L., Craven B., Utterback T.R., Khaliq H., Feldblyum T.V.,
RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
RL "Oryza sativa chromosome 10 BAC OSJNBa0003019 genomic sequence.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC060755; AAK00433.1; -.
DR Gramene; O9AYM0; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR000637; AT hook.
DR InterPro; IPR005175; DUF296.
DR Pfam; PF02178; AT hook; 2.
DR Pfam; PF03479; DUF296; 1.
DR SMART; SM00384; AT_hook; 2.
KW DNA-binding.
SQ SEQUENCE 405 AA; 41399 MW; 274994FD12D49816 CRC64;

Query Match 45.5%; Score 46; DB 10; Length 405;
Best Local Similarity 64.3%; Pred. No. 65;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 7 LALSGDGRVGSQC 20
:|:|:|:|:|:|
Db 272 VALAGSDGRVGGC 285

RESULT 10
Q7X976 PRELIMINARY; PRT; 405 AA.
AC Q7X976;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative AT-hook DNA-binding protein.
GN OSJNBa0003019.1.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA The Rice Chromosome 10 Sequencing Consortium;
RT "In-depth view of structure, activity, and evolution of rice
RL chromosome 10.";
RL Science 300:1566-1569(2003).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Buell C.R., Wing R.A., McCombie W.R., Messing J., Yuan Q.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE017121; AAP55117.1; -.
KW DNA-binding.
SQ SEQUENCE 405 AA; 41399 MW; 274994FD12D49816 CRC64;

Query Match 45.5%; Score 46; DB 10; Length 405;
Best Local Similarity 64.3%; Pred. No. 65;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

```

QY 7 LALSGFDGRVSGC 20
 :|||:|||||
 Db 272 VALAGSDGRVLGCG 285

RESULT 11

Q916M2 PRELIMINARY; PRT; 473 AA.
 AC Q916M2;
 DT 01-WAR-2001 (TReMBLrel. 16, Created)
 DT 01-WAR-2001 (TReMBLrel. 16, Last sequence update)
 DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
 DE Probable transcriptional regulator.
 GN PA0268.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Coltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RT opportunistic pathogen."
 RL Nature 406:959-964 (2000).
 CC -1- SIMILARITY: BELONGS TO THE GNTN FAMILY OF TRANSCRIPTIONAL
 CC REGULATORS.
 DR EMBL; AE004465; RAG03657.1; --
 DR PIR; D83611; D83611.
 DR GO; GO:0005622; C:intracellular; IEA.
 DR GO; GO:0003700; F:transcription factor activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR000524; HTH_GntR.
 DR Pfam; PF00392; gntR; 1.
 DR PRINTS; PR00035; HTHGNTR.
 DR SMART; SM00345; HTH_GNTR; 1.
 DR PROSITE; PS00043; HTH_GNTR_FAMILY; 1.
 KW DNA-binding; Transcription regulation; Complete proteome.
 SQ SEQUENCE 473 AA; 53313 MW; 3C6E9543B3D63F95 CRC64;

Query Match 45.0%; Score 45.5; DB 16; Length 473;
 Best Local Similarity 73.3%; Pred. No. 92;
 Matches 11; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 6 RLALSGFDGRVSGC 19
 :|||:|||||
 Db 59 RLASEGFLGRVSGC 73

RESULT 12

Q8EHA4 PRELIMINARY; PRT; 1482 AA.
 AC Q8EHA4;
 DT 01-WAR-2003 (TReMBLrel. 23, Created)
 DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
 DE Glutamate synthase, large subunit.
 GN GLTB OR SO1325.
 OS Shewanella oneidensis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
 OC Alteromonadaceae; Shewanella.
 OX NCBI_TaxID=70863;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NR-1;
 RX MEDLINE=22297686; PubMed=12368813;

RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
 RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
 RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
 RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
 RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
 RA Mueller J., Khouri H., Gill J., Uterback T.R., McDonald L.A., C.M.;
 RA Feldblyum T.V., Smith H.O., Venter J.C., Nealon K.H., Fraser C.M.;
 RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
 RT Shewanella oneidensis."
 RL Nat. Biotechnol. 20:1118-1123 (2002).
 DR EMBL; AE015576; AAN54390.1; --
 DR TIGR; SO1325; --
 DR GO; GO:0015930; F:glutamate synthase activity; IEA.
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.
 DR GO; GO:0006537; P:glutamate biosynthesis; IEA.
 DR GO; GO:0006807; P:nitrogen metabolism; IEA.
 DR InterPro; IPR002198; ADH_short.
 DR InterPro; IPR002489; DUF14.
 DR InterPro; IPR003009; FMN enzyme.
 DR InterPro; IPR002932; Glu_synthase.
 DR InterPro; IPR006982; Glu_synth_cent.
 DR InterPro; IPR006981; Glu_synth_NTN.
 DR Pfam; PF01645; Glu_synthase; 1.
 DR Pfam; PF04897; Glu_synth_NTN; 1.
 DR Pfam; PF04898; Glu_syn_central; 1.
 DR Pfam; PF01493; GXGXG; 1.
 DR PROSITE; PS00061; ADH_SHORT; 1.
 KW Complete proteome.
 SQ SEQUENCE 1482 AA; 163011 MW; 10FDB14B568BF474 CRC64;

Query Match 45.0%; Score 45.5; DB 16; Length 1482;
 Best Local Similarity 37.9%; Pred. No. 3.2e+02;
 Matches 11; Conservative 1; Mismatches 6; Indels 11; Gaps 1;

QY 3 RAVRLALSGF-----DGRVSGC 20
 :|||:|
 Db 28 RIVRTAIHGLDRMKRGRIASDGRGTGDC 56

RESULT 13

Q9LKR6 PRELIMINARY; PRT; 453 AA.
 ID Q9LKR6;
 AC Q9LKR6;
 DT 01-OCT-2000 (TReMBLrel. 15, Created)
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
 DT 01-WAR-2002 (TReMBLrel. 20, Last annotation update)
 DE T26D3.3 protein.
 GN T26D3.3.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA WashU;
 RT "The A. thaliana Genome Sequencing Project."
 RL Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Wilson R.;
 RL Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AF262043; AAF88016.1; --
 DR HSSP; P18670; 1JAC.
 DR InterPro; IPR001229; Jacalin_lectin.
 DR Pfam; PF01419; Jacalin; 3.
 SQ SEQUENCE 453 AA; 50205 MW; D37A33D35371E172 CRC64;

Query Match 44.6%; Score 45; DB 10; Length 453;
 Best Local Similarity 75.0%; Pred. No. 1.1e+02;

RC STRAIN=AX4;
RA Baumgart C.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC117075; AAC050777.1; -
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004356; F:glutamate-ammonia ligase activity; IEA.
DR GO; GO:0016874; F:ligase activity; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0003999; P:nitrogen fixation; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR008147; Gln_synt_beta.
DR InterPro; IPR008146; Gln_synt_C.
DR InterPro; IPR000508; Peptidase_S26.
DR Pfam; PF00120; gln-synt; 1.
DR ProDom; PD001057; Gln_synt_C; 1.
DR PROSITE; PS00761; SPASE_I_3; 1.
DR KW
DR LGase.
SQ SEQUENCE 499 AA; 56341 MW; 42A903AE0AA6365 CRC64;

Query Match 44.6%; Score 45; DB 5; Length 499;
Best Local Similarity 88.9%; Pred. NO. 1.2e+02;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 12 FDGRVSGC 20
||| |||||
DB 293 FDGLVSGC 301

RESULT 16

Q8UBB2 PRELIMINARY; PRT; 503 AA.

ID Q8UBB2 AC Q8UBB2
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Alpha-L-arabinofuranosidase.
DE ATU3104 OR AGR_L 3408.
GN Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OS Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;
[1]
SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RA "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58.";
RT Science 294:2317-2323 (2001).
[2]
SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmlet K., Gordon J., Vaudin M., Iarchouk O., Epp A., Liu F.,
RA Wollan C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58.";
RL Science 294:2323-2328 (2001).
DR EMBL; AE009241; AAL43920.1; -
DR EMBL; AE008372; AAK90279.1; -
DR PIR; AB2938; AB2938.
DR PIR; E98344; E98344.

```

KW Complete proteome.
SQ SEQUENCE 503 AA; 56495 MW; 3514F976D1854748 CRC64;

Query Match
Best Local Similarity 44.6%; Score 45; DB 16; Length 503;
Matches 8; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 5 VRLASGFGDVRGSG 19
   :|:|:|:|:|:|
Db 450 LHLAMAGYDLRVGNG 464

RESULT 17
Q7VGM3 PRELIMINARY; PRT; 408 AA.
AC Q7VGM3;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25; Last sequence update)
DE Hypothetical protein.
GN HH1298.
OS Helicobacter hepaticus.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Helicobacter.
OX NCBI_TaxID=32025;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 51449 / 3B1;
RX MEDLINE=22709201; PubMed=12810954;
RA Suerbaum S., Josenhans C., Stenzenbach T., Drescher B., Brandt P.,
RA Bell M., Droegge M., Farkmann B., Fischer H.-P., Ge Z., Hoerster A.,
RA Holland R., Klein K., Koenig J., Macko L., Mendz G.L., Nyakatura G.,
RA Schauer D.B., Shen Z., Weber J., Frosch M., Fox J.G.;
RT "The complete genome sequence of the carcinogenic bacterium
RT Helicobacter hepaticus."
RL Proc. Natl. Acad. Sci. U.S.A. 100:7901-7906(2003).
DR EMBL; AF017148; AAP77895.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 408 AA; 44657 MW; 7CEA58F77F875870 CRC64;

Query Match
Best Local Similarity 44.1%; Score 44.5; DB 16; Length 408;
Matches 9; Conservative 5; Mismatches 5; Indels 7; Gaps 1;

QY 2 QRAVRLA-----LSGFGDVRGSG 20
   :|:|:|:|:|:|:|:|
Db 106 QSNIRIAQLIGGIAGFGARIGMGC 131

RESULT 18
Q9UHE2 PRELIMINARY; PRT; 126 AA.
AC Q9UHE2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE RAD52 beta isoform (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Peripheral blood;
RX MEDLINE=20135586; PubMed=10673031;
RA Kito K., Wada H., Yeh E.T.H., Kamitani T.;
RT "Identification of novel isoforms of human RAD52."
RL Biochim. Biophys. Acta 1489:303-314(1999).
DR EMBL; AF187983; AAP05532.1; -.
DR InterPro; IPR007232; Rad52_Rad22.
DR Pfam; PF04098; Rad52_Rad22; 1.
FT NON_TER 1
FT NON_TER 126

```

```

SQ SEQUENCE 126 AA; 13661 MW; 5C1F0F4B02DD061F CRC64;

Query Match
Best Local Similarity 43.6%; Score 44; DB 4; Length 126;
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 LQRAVRLALSGFGDVRV 16
   :|:|:|:|:|:|
Db 64 LKRALRLPLGVSGRI 79

RESULT 19
Q98J09 PRELIMINARY; PRT; 136 AA.
AC Q98J09;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein mlr2160.
GN MLR2160.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idegawa K., Ishikawa K., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti."
RL DNA Res. 7:331-338(2000).
DR EMBL; AP002999; BAB49357.1; -.
DR GO; GO:0003824; P:catalytic activity; IEA.
DR GO; GO:0006633; P:fatty acid biosynthesis; IEA.
DR InterPro; IPR007272; DUF395.
DR InterPro; IPR000794; Ketoacyl_synth.
DR Pfam; PF04143; DUF395; 1.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 136 AA; 13645 MW; 7F3C7C9667800550 CRC64;

Query Match
Best Local Similarity 43.6%; Score 44; DB 16; Length 136;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 9 LSGFGDVRGSGC 20
   :|:|:|:|:|:|
Db 86 LVGFGSRWGGC 97

RESULT 20
Q92ZC0 PRELIMINARY; PRT; 136 AA.
AC Q92ZC0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein RA0570.
GN RA0570 OR SMA1053.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OG Plasmid pSMA (megaplasmid 1).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396509; PubMed=11481432;

```

RESULT	22
Q96GW2	
ID	PRELIMINARY; PRT; 485 AA.
AC	Q96GW2;
DT	01-DEC-2001 (TrEMBLrel. 19, Created)
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE	Hypothetical protein.
DE	Homo sapiens (Human).
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

RESULT 24	
Q93YH6	
ID Q93YH6	PRELIMINARY; PRT; 414 AA.
AC Q93YH6;	
DT 01-DEC-2001	(TREMBLrel. 19, Created)
DT 01-DEC-2001	(TREMBLrel. 19, Last sequence update)
DT 01-OCT-2003	(TREMBLrel. 25, Last annotation update)
DT 01-OCT-2003	(TREMBLrel. 25, Last annotation update)

```

DE Glyceraldehyde-3-phosphate dehydrogenase (NADP+) (EC 1.2.1.13).
OS Galdieria sulphuraria (Red alga).
OC Eukaryota; Rhodophyta; Bangiophyceae; Porphyridiales; Porphyridiaceae;
OC Galdieria.
OX NCBI_TaxID=130081;
RN [1]_TaxID=130081;
RP SEQUENCE FROM N.A.
RA Herrig M., Tegeler A., Scheibe R.;
RT "Identification of a GAPDH-encoding cDNA from Galdieria sulphuraria,
RT heterologous expression and characterization of the enzyme.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE
CC DEHYDROGENASE FAMILY.
DR EMBL: AJ012286; CAC80066.1; -.
DR HSSP: P00354; 3GPD.
DR GO: GO:0009507; C:chloroplast; IEA.
DR GO: GO:0047100; F:glyceraldehyde-3-phosphate dehydrogenase (N. . . ; IEA.
DR GO: GO:004365; F:glyceraldehyde-3-phosphate dehydrogenase (p. . . ; IEA.
DR GO: GO:0016491; F:oxidoreductase activity; IEA.
DR GO: GO:0006096; P:glycolysis; IEA.
DR InterPro: IPR006424; GAPDH-I.
DR Pfam: PF00044; gpdh; 1.
DR Pfam: PF02800; gpdh; C; 1.
DR PRINTS: PR00078; G3PDHGRNASE.
DR TIGRFAMs: TIGR01534; GAPDH-I; 1.
DR PROSITE: PS00071; GAPDH; 1.
KW NADP; Oxidoreductase; Transit peptide; Chloroplast.
FT TRANSIT 1 77 CHLOROPLAST
FT CHAIN 78 414 GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE
FT SEQUENCE 414 AA; 45446 MW; 42E566EB32E14A9D CRC64;

Query Match 43.1%; Score 43.5; DB 10; Length 414;
Best Local Similarity 52.9%; Pred. No. 1.7e+02;
Matches 9; Conservative 5; Mismatches 2; Indels 1; Gaps 1;

QY 1 LQRAVRLALSGFDGRVG 17
DB 75 MQAKVRVAINGF-GRIG 90

RESULT 25
Q8PHP2 PRELIMINARY; PRT; 596 AA.
AC Q8PHP2;
DT 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DE Hypothetical protein XAC3208.
GN XAC3208.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
RA Da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardoso J., Chambergo F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.C., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,

Setubal J.C., Kitajima J.P.;
"Comparison of the genomes of two Xanthomonas pathogens with differing
host specificities.";
RL Nature 417:459-463 (2002).
DR EMBL: AE011966; BAM38051.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 596 AA; 64542 MW; 40B880665884FCD7 CRC64;

Query Match 43.1%; Score 43.5; DB 16; Length 596;
Best Local Similarity 60.0%; Pred. No. 2.5e+02;
Matches 12; Conservative 1; Mismatches 6; Indels 1; Gaps 1;

QY 1 LQRAVRLALSGFDGRVSGC 20
DB 27 LQRAARSA-SGRGRKGADC 45

RESULT 26
Q82IR7 PRELIMINARY; PRT; 147 AA.
AC Q82IR7;
DT 01-JUN-2003 (T-EMBLrel. 24, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN SAV3066.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220 (2001).
RN [2]

SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis.";
RL Nat. Biotechnol. 21:526-531 (2003).
DR EMBL: AP005033; BAC70777.1; -.
DR GO: GO:0008237; P:metalloproteinase activity; IEA.
DR GO: GO:0008270; F:zinc ion binding; IEA.
DR GO: GO:0008508; P:proteolysis and peptidolysis; IEA.
DR InterPro: IPR006025; Pept M Zn BS.
DR PROSITE: PS00142; ZINC PROTEASE; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 147 AA; 16784 MW; 88516363F70A09DD CRC64;

Query Match 42.6%; Score 43; DB 16; Length 147;
Best Local Similarity 64.3%; Pred. No. 65;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 QRAVRLALSGFDGR 15
DB 39 ERAERLSQYFDR 52

RESULT 27
Q81SJ0 PRELIMINARY; PRT; 254 AA.
ID Q81SJ0
AC Q81SJ0;
DT 01-JUN-2003 (T-EMBLrel. 24, Created)

```

```

DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Chomataxis protein MotA, putative.
GN BAI658.
OS Bacillus anthracis (strain Ames).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=198094;
RN [1]_TaxID=198094;
RP SEQUENCE FROM N.A.
RX MEDLINE=22608414; PubMed=12721629;
RA Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
RA Holtzapple E.K., Okstad O.A., Helgason E., Rilstone J., Wu M.,
RA Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,
RA DeBoy R.T., Madpu R., Daugherty S.C., Duskin A.S., Haft D.H.,
RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
RA Benton J.L., Mahmoud Y., Jiang L., Hance I.R., Weidman J.F.,
RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
RA Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,
RA Thomson B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,
RA Fraser C.M.;
RT "The genome sequence of Bacillus anthracis Ames and comparison to
RT closely related bacteria."
RL Nature 423:81-86(2003).
DR EMBL; AE017029; AAP225591.1; -.
DR TIGR; BAI658; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0003774; F:motor activity; IEA.
DR GO; GO:0008565; F:protein transporter activity; IEA.
DR GO; GO:0006935; P:chemotaxis; IEA.
DR GO; GO:0001539; P:ciliary/flagellar motility; IEA.
DR GO; GO:0005810; P:transport; IEA.
DR InterPro; IPR000540; Flag_MotA.
DR InterPro; IPR002898; MotA_Exbb.
DR Pfam; PF01618; MotA_Exbb; 1.
DR PROSITE; PS01307; MOT_A; 1.
KW Complete proteome.
SQ SEQUENCE 254 AA; 28179 MW; 5FA623C72B7E1D5F CRC64;

Query Match 42.6%; Score 43; DB 16; Length 254;
Best Local Similarity 53.8%; Pred. No. 1.2e+02;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 LQRAVRLALSGPD 13
Db 105 IQKGIRLMLSGYD 117

RESULT 28
O81FG5 PRELIMINARY; PRT; 254 AA.
AC Q81FG5;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DE Chomataxis motA protein.
DE Chomataxis motA protein.
GN BC1625.
OS Bacillus cereus (strain ATCC 14579 / DSM 31).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=226900;
RN [1]_TaxID=226900;
RP SEQUENCE FROM N.A.
RX MEDLINE=22608415; PubMed=12721630;
RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,
RA Kapural V., Bhattacharyya A., Reznik G., Mikhailova N., Lapidus A.,
RA Chu L., Mazur M., Goltsman E., Larsen N., D'Souza M., Walunas T.,
RA Grechkin Y., Pusch G., Haselkorn R., Fonstein M., Ehrlich S.D.,
RA Overbeek R., Kyrpides N.;
RT "Genome sequence of Bacillus cereus and comparative analysis with
RT Bacillus anthracis."
RL Nature 423:87-91(2003).
DR EMBL; AE017003; AAP08604.1; -.
DR GO; GO:0016020; C:membrane; IEA.

DR GO; GO:0003774; F:motor activity; IEA.
DR GO; GO:0008565; F:protein transporter activity; IEA.
DR GO; GO:0006935; P:chemotaxis; IEA.
DR GO; GO:0001539; P:ciliary/flagellar motility; IEA.
DR GO; GO:0005810; P:transport; IEA.
DR InterPro; IPR000540; Flag_MotA.
DR InterPro; IPR002898; MotA_Exbb.
DR Pfam; PF01618; MotA_Exbb; 1.
DR PROSITE; PS01307; MOT_A; 1.
KW Complete proteome.
SQ SEQUENCE 254 AA; 28179 MW; 5FA623C72B7E1D5F CRC64;

Query Match 42.6%; Score 43; DB 16; Length 254;
Best Local Similarity 53.8%; Pred. No. 1.2e+02;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 LQRAVRLALSGPD 13
Db 105 IQKGIRLMLSGYD 117

RESULT 28
O81FG5 PRELIMINARY; PRT; 254 AA.
AC Q81FG5;
DT 01-JUN-2003 (TReMBLrel. 24, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DE Chomataxis motA protein.
DE Chomataxis motA protein.
GN BC1625.
OS Bacillus cereus (strain ATCC 14579 / DSM 31).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=226900;
RN [1]_TaxID=226900;
RP SEQUENCE FROM N.A.
RX MEDLINE=22608415; PubMed=12721630;
RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,
RA Kapural V., Bhattacharyya A., Reznik G., Mikhailova N., Lapidus A.,
RA Chu L., Mazur M., Goltsman E., Larsen N., D'Souza M., Walunas T.,
RA Grechkin Y., Pusch G., Haselkorn R., Fonstein M., Ehrlich S.D.,
RA Overbeek R., Kyrpides N.;
RT "Genome sequence of Bacillus cereus and comparative analysis with
RT Bacillus anthracis."
RL Nature 423:87-91(2003).
DR EMBL; AE017003; AAP08604.1; -.
DR GO; GO:0016020; C:membrane; IEA.

DR GO; GO:0003774; F:motor activity; IEA.
DR GO; GO:0008565; F:protein transporter activity; IEA.
DR GO; GO:0006935; P:chemotaxis; IEA.
DR GO; GO:0001539; P:ciliary/flagellar motility; IEA.
DR GO; GO:0005810; P:transport; IEA.
DR InterPro; IPR000540; Flag_MotA.
DR InterPro; IPR002898; MotA_Exbb.
DR Pfam; PF01618; MotA_Exbb; 1.
DR PROSITE; PS01307; MOT_A; 1.
KW Complete proteome.
SQ SEQUENCE 254 AA; 28090 MW; 82227B95BB72CA5A CRC64;

Query Match 42.6%; Score 43; DB 16; Length 254;
Best Local Similarity 53.8%; Pred. No. 1.2e+02;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 LQRAVRLALSGPD 13
Db 105 IQKGIRLMLSGYD 117

RESULT 29
O83611 PRELIMINARY; PRT; 287 AA.
AC O83611;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Phosphatidate cytidyltransferase (CDSA).
GN TP0602.
OS Treponema pallidum.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Nichols;
RX MEDLINE=98332770; PubMed=9665876;
RA Dodson R.R., Norris S.J., Weinstein G.M., White O., Sutton G.G.,
RA Fraser C.M., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Saizberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete."
RL Science 281:375-388(1998).
DR EMBL; AE001235; AAC65575.1; -.
DR PIR; D71304; D71304.
DR TIGR; TP0602; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004605; P:phosphatidate cytidyltransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0008654; P:phospholipid biosynthesis; IEA.
DR InterPro; IPR000374; PC_trans.
DR Pfam; PF01148; CTP_transf_1.
KW Transferase; Nucleotidyltransferase; Complete proteome.
SQ SEQUENCE 287 AA; 30866 MW; 426D08D3A0C1B0FB CRC64;

Query Match 42.6%; Score 43; DB 16; Length 287;
Best Local Similarity 58.3%; Pred. No. 1.4e+02;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 9 LSGFGRVGSQC 20
Db 193 IGGFAGSVGAGC 204

RESULT 30
Q92AC6 PRELIMINARY; PRT; 352 AA.
ID Q92AC6
AC Q92AC6;
DT 01-DEC-2001 (TReMBLrel. 19, Created)

```


DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein lin1996.
GN LIN1996.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / Serovar 6a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rueniok C., Amend A.,
RA Baquero F., Berche P., Blöcker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,
RT "Comparative genomics of Listeria species."
RL Science 294:849-852(2001).
DR EMBL; AL596170; CAC97226.1; -.
DR PIR; AB1682; AB1682.
DR ListList; LIN01996; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001223; Glyco_hydro_18.
DR InterPro; IPR001579; Glyco_hydro_18AS.
DR Pfam; PF00704; Glyco_hydro_18; 1.
DR ProDom; PD000471; Glyco_hydro_18; 1.
DR PROSITE; PS01095; CHITINASE 18; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 352 AA; 37949 MW; 771A0DC781C11359 CRC64;

Query Match 42.6%; Score 43; DB 16; Length 352;
Best Local Similarity 71.4%; Pred. NO. 1.7e+02;
Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 RAVRLALSGFDGRV 16
DB 119 RAVLLALGGADGHV 132

Search completed: May 13, 2004, 06:55:15
Job time : 37.3333 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2004, 06:49:47 ; Search time 48 Seconds
(without alignments)
117.728 Million cell updates/sec

Title: 09549186-9

Perfect score: 99

Sequence: 1 LQRAVRLALSGSDRGVSGC 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : A_Geneseq_29Jan04:*

1: Geneseqp1980s:*

2: Geneseqp1990s:*

3: Geneseqp2000s:*

4: Geneseqp2001s:*

5: Geneseqp2002s:*

6: Geneseqp2003as:*

7: Geneseqp2003bs:*

8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	90	90.9	20	2	ADB87347 Vaccinati
2	68	68.7	19	2	AAR45910 Foot and
3	48	48.5	346	4	ABG06006 Novel hum
4	48	48.5	898	5	ABG70103 Human pre
5	48	48.5	1872	2	AAR56493 TATA-bind
6	48	48.5	1872	2	AAR06078 Drosophil
7	48	48.5	1872	2	AAR25030 TATA-bind
8	48	48.5	1893	2	AAR56491 TATA-bind
9	48	48.5	1893	2	AAR06082 Human TAR
10	48	48.5	1893	2	AAR25020 TATA-bind
11	48	48.5	1924	4	ABG06008 Novel hum
12	47	47.5	67	4	ABG43462 Propionib
13	47	47.5	67	6	ABM39981 Propionib
14	46	46.5	220	4	ABG29258 Novel hum
15	46	46.5	352	5	ABBA48949 Listeria
16	45	45.5	17	4	AAM00762 Human bon
17	45	45.5	359	4	AAR33620 Pseudomon
18	45	45.5	359	6	ABU15565 Protein e
19	45	45.5	394	4	ABBS59456 Xenopus h
20	44	44.4	73	2	AAY22514 Xenopus h
21	44	44.4	158	4	ABBG3810 Drosophil
22	44	43.4	133	7	ADC07742 Rice prot
23	43	43.4	215	3	AAG46917 Arabidops
24	43	43.4	233	6	ABM67849 Photorhab
25	43	43.4	258	3	AAG46916 Arabidops

26	43	43.4	297	3	AAG46915 Arabidops
27	43	43.4	349	4	ABG11319 Novel hum
28	43	43.4	436	6	ABM68505 Photorhab
29	43	43.4	585	2	AAR34261 An alpha
30	43	43.4	1017	4	AAR96815 Putative
31	42.5	42.9	229	4	AAG2858 C Glutami
32	42.5	42.9	986	6	ABU14884 Protein e
33	42	42.4	73	2	AAY22513 Human hnr
34	42	42.4	85	4	AAM22023 Peptide #
35	42	42.4	85	4	ABB44405 Peptide #
36	42	42.4	85	4	AAM38388 Peptide #
37	42	42.4	85	4	ABB27256 Protein #
38	42	42.4	85	4	AAM78147 Human bon
39	42	42.4	85	4	ABM65483 Human liv
40	42	42.4	85	4	ABG59794 Human liv
41	42	42.4	85	5	ABG47160 Human pep
42	42	42.4	129	3	AAG01219 Human sec
43	42	42.4	129	3	AAG01218 Human sec
44	42	42.4	132	4	ABG23137 Novel hum
45	42	42.4	135	3	ABBS8154 Lung canc
46	42	42.4	167	4	ABG22809 Novel hum
47	42	42.4	187	2	AAY60035 Human end
48	42	42.4	224	2	AAY60124 Human end
49	42	42.4	306	6	ABO52970 Human spl
50	42	42.4	310	4	AAU68591 Human nov
51	42	42.4	314	4	ABG03426 Novel hum
52	42	42.4	409	6	ABU44363 Protein e
53	42	42.4	511	4	AAB67687 Amino aci
54	42	42.4	547	3	AAY56849 Human RNA
55	42	42.4	547	4	AAB67685 Amino aci
56	42	42.4	547	4	AAM39379 Human pol
57	42	42.4	547	4	AAB95154 Human pro
58	42	42.4	547	4	ABU52800 Human nuc
59	42	42.4	547	6	ABU62208 Human RNA
60	42	42.4	547	7	ADB99123 Human RNA
61	42	42.4	548	2	AAR88549 Secreted
62	42	42.4	548	4	ABBS0316 Human sec
63	42	42.4	548	6	ABO44573 Human pro
64	42	42.4	548	7	ABO26053 Human pro
65	42	42.4	566	4	AAM41165 Human pol
66	42	42.4	596	4	ABSG7493 Drosophil
67	42	42.4	600	5	ABU65148 Human NOV
68	42	42.4	600	7	ADC31316 Human nov
69	42	42.4	610	4	ABG19757 Novel hum
70	42	42.4	639	4	ABBT70075 Drosophil
71	42	42.4	855	4	ABB71577 Drosophil
72	42	42.4	1477	2	AAB67691 S. cerevi
73	42	42.4	1477	2	AAW10424 Saccharom
74	42	42.4	1477	2	AAU06819 Fumonosin
75	42	42.4	1477	6	ABRS3665 Protein s
76	42	42.4	1477	7	ABBS0301 Human Y-b
77	42	42.4	19938	6	ABBS98398 Streptomy
78	41.5	41.9	78	4	AAM92483 Human dig
79	41.5	41.9	78	4	AAU22528 Novel hum
80	41.5	41.9	78	7	ADB32368 Human nov
81	41.5	41.9	787	6	ABU17373 Protein e
82	41.5	41.9	1485	5	ABU41173 Protein e
83	41	41.4	43	5	AAU83528 Novel hum
84	41	41.4	84	4	ABG13507 Novel hum
85	41	41.4	89	4	ABBS98897 Respirato
86	41	41.4	141	4	AAU63931 Propionib
87	41	41.4	141	6	ABM60450 Propionib
88	41	41.4	147	3	AAG41096 Zea maye
89	41	41.4	162	7	ADD28290 Human het
90	41	41.4	166	7	ADD28254 Human het
91	41	41.4	169	4	AAU64886 Propionib
92	41	41.4	169	6	ABM61405 Propionib
93	41	41.4	179	3	AAG41095 Zea maye
94	41	41.4	240	2	AAR78738 Murine mV
95	41	41.4	240	2	AAW26594 Murine bM
96	41	41.4	240	4	AAE10984 Murine pa
97	41	41.4	240	4	ABG73300 Amino aci
98	41	41.4	244	4	AAU54130 Propionib

99 41 41.4 244 6 ABM50649 Propionib
100 41 41.4 269 6 ABU49525 Protein e

ALIGNMENTS

```
RESULT 1
ADB87347
XX ADB87347 standard; peptide; 20 AA.
AC ADB87347;
XX
XX 04-DEC-2003 (first entry)
DE Vaccination related retro-partly inverso peptide #1.
XX
KW immunoretroid; anti-immunoretroid; CONH linkage; NHCO linkage;
KW retropeptide; retroinverso peptide; vaccine; viral; bacterial infection;
KW autoimmune disease; neurodegenerative disease; retro-partly;
KW inverso peptide.
XX
OS Unidentified.
XX
FH Key Location/Qualifiers
FT Modified-site 1 /note= "Modified by OH-m. Peptide can be an R or S
FT isomer"
FT Misc-difference 2 /note= "D-form residue"
FT Misc-difference 3 /note= "D-form residue"
FT Misc-difference 4 /note= "D-form residue"
FT Misc-difference 5 /note= "D-form residue"
FT Misc-difference 6 /note= "D-form residue"
FT Misc-difference 7 /note= "D-form residue"
FT Misc-difference 8 /note= "D-form residue"
FT Misc-difference 9 /note= "D-form residue"
FT Misc-difference 10 /note= "D-form residue"
FT Misc-difference 11 /note= "D-form residue"
FT Misc-difference 12 /note= "D-form residue"
FT Misc-difference 13 /note= "D-form residue"
FT Misc-difference 14 /note= "D-form residue"
FT Misc-difference 15 /note= "D-form residue"
FT Misc-difference 16 /note= "D-form residue"
FT Misc-difference 17 /note= "D-form residue"
FT Modified-site 18 /note= "C-terminal amide"
XX
XX FR2717081-A1.
XX
XX 15-SEP-1995.
XX
XX 14-MAR-1994; 94FR-00002950.
XX
XX 14-MAR-1994; 94FR-00002950.
XX
XX (CNRS ) CENT NAT RECH SCI.
XX
XX Guichard G, Muller S, Briand J, Regenmortel MHV;
```

```
DR WPI; 1995-322414/42.
XX
PT Therapeutic and diagnostic uses of retro peptide analogues - corresp. to
PT parent peptide chains with CONH linkages replaced by NHCO linkages, also
PT antibodies against the peptide(s).
XX
PS Disclosure; Page 21; 58pp; French.
XX
XX This invention relates to the novel uses of 'immunoretroids' or anti-
XX immunoretroid antibodies, where the immunoretroids are peptide analogues
XX in which one or more (preferably all) of the CONH linkages in the chain
XX of the corresponding parent peptides are replaced by NHCO linkages and
XX the chirality of each amino acid residue, whether involved in NHCO
XX linkages or not, is either conserved or inverted with regards to the
XX corresponding amino acid residue in the parent peptides. For example,
XX 'retropeptides' or 'retroinverso peptides', provided that the
XX immunoretroids are capable of forming complexes with the anti-
XX immunoretroid antibodies and with antibodies directed against the parent
XX peptides or parent proteins and/or the parent peptide enantiomers or
XX parent protein enantiomers. The immunoretroids are used to prepare
XX medicaments for preventing or treating pathologies associated with the
XX presence of an exogenous or endogenous protein capable of being
XX implicated directly or indirectly in the appearance and/or development of
XX the pathologies. Immunoretroids can also be used to prepare vaccines for
XX preventing pathologies associated with the presence of an exogenous or
XX endogenous protein recognised by antibodies directed against
XX immunoretroids. Comparisons containing immunoretroids associated with a
XX carrier molecule capable of inducing production of antibodies against an
XX exogenous or endogenous protein responsible for a pathology, or of
XX inducing a cytotoxic cellular immune response are useful as vaccines.
XX Pathologies that can be diagnosed or treated are especially viral or
XX bacterial infections, autoimmune diseases and neurodegenerative diseases.
XX This sequence represents a vaccination related retro-partly inverso
XX peptide relating to the retropeptides of the invention.
XX
SQ Sequence 20 AA;
Query Match 90.9%; Score 90; DB 2; Length 20;
Best Local Similarity 90.0%; Pred. No. 1.8e-07;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 LQRAVRLALSGSDGRVSGSC 20
Db 1 LQRAVRLALSGSDGRVSGSC 20
||||| ||||| ||||| |||||
RESULT 2
AAR45910
ID AAR45910 standard; peptide; 19 AA.
XX
XX AAR45910;
XX
XX 27-AUG-2003 (revised)
DT 25-MAR-2003 (revised)
DT 27-JUL-1994 (first entry)
XX
DE Foot and Mouth Disease Virus immunogenic peptide P6.
XX
KW immunodominant surface-exposed epitope; T-cell proliferation;
KW stimulation; immunogenicity; increase; enhance; vaccine;
KW Foot and Mouth Disease Virus; FMDV; Aphthovirus; LCP; lipid anchor;
KW immunogenic carrier.
XX
XX Foot-and-mouth disease virus.
OS
XX
XX WO9402506-A1.
PN
XX
XX 03-FEB-1994.
PD
XX
XX 23-JUL-1993; 93WO-GB001558.
PF
XX
XX 24-JUL-1992; 92GB-00015780.
PR
XX
```

PA (UNLO) UNIV LONDON SCHOOL PHARMACY.
 PI Toth I, Gibbons WA;
 XX
 XX
 DR WPI; 1994-048791/06.
 XX
 PT New lipidic amino acid based anchor system - for attachment of short
 PT synthetic peptide(s) to enhance their antigenicity and for use, e.g., as
 PT vaccines.
 XX
 XX
 PS Example 3; Page 36; 50pp; English.
 XX
 CC A lipidic amino acid based anchor system was synthesised. Eight copies of
 CC the FMDV immunogenic peptide P6 (AAR45910) were attached to the anchor to
 CC give (peptide)8LYs4LYs2LYs(HNCH(CH2)13ME)CO3NH2. When injected into
 CC cows, immunogenicity was found to be 10 times higher than would be
 CC expected from a conventional BSA- or KLH-conjugated FMDV. (Updated on 25-
 CC MAR-2003 to correct PN field.) (Updated on 27-AUG-2003 to correct OS
 CC field.)
 XX
 SQ Sequence 19 AA;
 Query Match 68.7%; Score 68; DB 2; Length 19;
 Best Local Similarity 88.2%; Pred. No. 0.0062; Mismatches 2; Indels 0; Gaps 0;
 Matches 15; Conservative 0;
 QY 3 RAVRLALSGDGRVGS 19
 ||||| ||||| |||||
 DB 3 RAVRPALSGFGRVGS 19
 ||||| ||||| |||||
 RESULT 3
 ABG06006
 ID ABG06006 standard; protein; 346 AA.
 XX
 AC ABG06006;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #5997.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US008631.
 XX
 PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS70193.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 20; SEQ ID NO 36365; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used

CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG03077 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 346 AA;
 Query Match 48.5%; Score 48; DB 4; Length 346;
 Best Local Similarity 62.5%; Pred. No. 28;
 Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 4 AVRLALSGDGRVGS 19
 ||||| ||||| |||||
 DB 214 AQLSESGSDSVGS 229
 ||||| ||||| |||||
 RESULT 4
 ABG70103
 ID ABG70103 standard; protein; 898 AA.
 XX
 AC ABG70103;
 XX
 DT 21-OCT-2002 (first entry)
 XX
 DE Human prey protein for Shigella ospC1 #21.
 XX
 KW prey protein; ospB; ospD1; ipaC; ipaH9.8; ospG; ospC1; Shigella;
 KW shigellosis; bacillary dysentery; antibacterial; yeast two-hybrid system;
 KW protein-protein interaction; SID; selected interacting domain; human.
 XX
 OS Homo sapiens.
 XX
 PN WO200257303-A2.
 XX
 PD 25-JUL-2002.
 XX
 PF 11-JAN-2002; 2002WO-EP000777.
 XX
 PR 12-JAN-2001; 2001US-0261130P.
 XX
 PA (HYBR-) HYBRIGENICS.
 XX
 PI Legrain P;
 XX
 DR WPI; 2002-599706/64.
 DR N-PSDB; ABS51496.
 XX
 PT New complex of protein-protein interactions between a bait Shigella
 PT flexneri polypeptide and a prey mammalian or human placenta polypeptide
 PT for treating or preventing bacillary dysentery in a mammal or human.
 XX
 PS Claim 7; Page 80-81; 162pp; English.
 XX
 CC The invention relates to a complex of protein-protein interactions
 CC between a Shigella flexneri polypeptide (e.g. ospB, ospD1, ipaD, ipaC,
 CC ipaH9.8, ospG and ospC1) and a mammalian polypeptide defined in the
 CC specification. The complexes are formed using the yeast two-hybrid
 CC system. Also included are (1) a recombinant host cell expressing the
 CC interactions between the Shigella flexneri polypeptide and a mammalian
 CC polypeptide defined in the specification; (2) selecting a modulating

CC compound that inhibits or activates the protein-protein interactions; (3)
 CC a modulating compound obtained from the method of (2); (4) a SID
 CC (selected interacting domain) polypeptide or its fragment or variant
 CC comprising the human polypeptides appearing as ABG70042-ABG70242; (5) a
 CC SID polynucleotide or its fragment or variant comprising encoding the
 CC above polypeptides a vector comprising (5); (6) a recombinant host cell
 CC containing the vector; and (10) a protein chip comprising Shigella
 CC flexneri polypeptide and a mammalian polypeptide defined in the
 CC specification. A pharmaceutical composition comprising the compound,
 CC polypeptide or polynucleotide is useful for treating or preventing
 CC shigellosis (bacillary dysentery) in a human or mammal. The present
 CC sequence represents a human prey protein isolated by the yeast two-hybrid
 CC assay, forming a complex of the invention with a shigella protein
 XX
 XX SQ Sequence 898 AA;

Query Match 48.5%; Score 48; DB 5; Length 898;
 Best Local Similarity 62.5%; Pred. No. 82;
 Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 AVRLALSGSDGRVSG 19
 |:::| | | | |
 DB 766 AIQLSESGSDSDVSG 781

RESULT 5

AA056493
 ID AAR56493 standard; protein; 1872 AA.

XX AAR56493;

XX 25-MAR-2003 (revised)

DT 23-MAR-1995 (first entry)

XX TATA-binding protein-associated factor dTAFII250.

KW TATA-binding protein associated factor; dTAFII250; screening; diagnostic;
 KW therapeutic; gene transcription regulation.

XX Drosophila.

XX WO9417087-A1.

XX 04-AUG-1994.

PF 28-JAN-1994; 94WO-US001114.

XX 28-JAN-1993; 93US-00013412.

PR 30-JUN-1993; 93US-00087119.

XX (REGC) UNIV CALIFORNIA.

XX Tjian R, Comai L, Dynlacht BD, Hoey T, Ruppert S, Tanese N;

PI Wang E, Weinzierl ROJ;

XX WPI; 1994-264019/32.

DR N-PSDB; AAQ70725.

XX TATA-binding protein associated protein factors - and corresponding
 PT nucleotide sequence and deriv. antibodies, useful in screening,
 PT diagnostics and therapeutics.

XX Disclosure; Page 136; 180pp; English.

XX N.B. This protein sequence does not correspond to any of the 3 reading
 CC frames of the DNA sequence AAQ70725 (dTAFII250). The TATA-binding protein
 CC associated factor dTAFII250 (including specific antibodies and fusion
 CC products) are used in drug screening, diagnostics and therapeutics. They
 CC are used in the development of specific biochemical assays for screening
 CC compounds that agonise or antagonise selected transcription factors
 CC involved in regulating gene expression associated with human pathology.
 CC (Updated on 25-MAR-2003 to correct PN field.)
 XX

SQ Sequence 1872 AA;

Query Match 48.5%; Score 48; DB 2; Length 1872;
 Best Local Similarity 62.5%; Pred. No. 1.9e+02;
 Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 AVRLALSGSDGRVSG 19
 |:::| | | | |
 DB 1741 AIQLSESGSDSDVSG 1756

RESULT 6

AAW06078

ID AAW06078 standard; protein; 1872 AA.

XX AAW06078;

XX 25-MAR-2003 (revised)

DT 27-JAN-1997 (first entry)

XX Drosophila TATA-binding protein associated factor dTAFII250 protein.

KW Drosophila; TATA-binding protein; TBP associated factor; TFIID;

KW RNA polymerase II; transcription; messenger RNA; nuclear fraction;

KW holoenzyme; lambda-gt11; expression library.

XX Drosophila melanogaster.

XX US534410-A.

XX 09-JUL-1996.

PF 28-JAN-1994; 94US-00188582.

XX 28-JAN-1993; 93US-00013412.

PR 30-JUN-1993; 93US-00087119.

XX (REGC) UNIV CALIFORNIA.

XX Comai L, Hoey T, Tanese N, Ruppert S, Weinzierl ROJ, Tjian R;

PI Wang E, Dynlacht BD;

XX WPI; 1996-333245/33.

XX Screen for cpds. that bind human TATA-binding protein associated factor -
 PT by testing ability to bind to polypeptide fragments of the factor, useful
 PT as (ant)agonists of transcription factors involved in disease.

XX Example; Col 95-104; 86pp; English.

XX This is the amino acid sequence of the Drosophila TATA-binding protein
 CC (TBP) associated factor (TAF) designated TAFII250. The protein is a
 CC component of the TFIID fraction required for reconstituting RNA
 CC polymerase II in vitro transcription activity. The gene encoding the
 CC protein was isolated from a lambda-gt11 expression library prepared from
 CC 6-12 hr old embryos, using the monoclonal antibodies 2B2 and 30H9. A
 CC fragment of the TAFII250 gene mapped the TAFII250 gene to position 32E1-2
 CC on the left arm of chromosome 2. The invention relates to purified
 CC proteins involved in transcription by RNA polymerase II, the RNA
 CC polymerase which transcribes messenger RNA. RNA polymerase II
 CC transcription proceeds in vitro upon addition of several nuclear
 CC fractions designated TFIIA, B, D, E, F, H, I and J to RNA polymerase II
 CC holoenzyme designated TFIID. TFIID has been shown to contain a TBP and other
 CC TAFs. Purification of TFIID and separation of its components reveals 7
 CC proteins ranging in size from 30-250 kD. Serum raised against the TFIID
 CC fraction allowed cloning of the corresp. genes from lambda-gt11
 CC expression libraries. (Updated on 25-MAR-2003 to correct PF field.)
 XX

SQ Sequence 1872 AA;

Query Match 48.5%; Score 48; DB 2; Length 1872;
 Best Local Similarity 62.5%; Pred. No. 1.9e+02;
 Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 AVRLLALSGSDGRVGS 19
 DT :::: |||| ||||
 DB 1740 AIQLSESGSDSDVGS 1755

RESULT 7
 AAW25030
 ID AAW25030 standard; protein; 1872 AA.
 AC AAW25030;
 XX
 DT 25-MAR-2003 (revised)
 DT 08-OCT-1997 (first entry)
 XX
 DE TATA-binding protein associated factor, dTAFII250.
 XX
 KW TATA-binding protein associated factor; TAF; nuclear protein;
 KW RNA polymerase transcription; TATA-binding protein; TBP; initiation.
 XX
 OS Drosophila sp.
 XX
 PN US5637686-A.
 XX
 PD 10-JUN-1997.
 XX
 PF 09-MAY-1996; 96US-00646715.
 XX
 PR 28-JAN-1993; 93US-00013412.
 PR 30-JUN-1993; 93US-00087119.
 PR 28-JAN-1994; 94US-00188582.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Hoey T, Dynlact BD, Tjian R, Wang E, Weinzierl ROJ, Ruppert S;
 PI Tanese N, Comai L;
 PI
 DR WPI; 1997-319113/29.
 XX
 PT Nucleic acids encoding human TATA-binding protein associated factor (TAF)
 PT peptide(s) - for production of recombinant peptide(s), used for
 PT modulating transcription of TAFs.
 XX
 PS Example 1; Col 103-106; 86pp; English.
 XX
 CC AAW25030 represents TATA-binding protein associated factor (TAF)
 CC polypeptide, dTAFII250 (mol. weight 250kD). TAF peptides derived from
 CC dTAFII30 alpha, dTAFII30 beta, dTAFII40, dTAFII60, dTAFII80, dTAFII10,
 CC dTAFII150, and dTAFII250, their human equivalents and nucleic acids
 CC encoding them, are used to modulate transcription, including
 CC transcription initiation. TAFs are nuclear proteins involved in RNA
 CC polymerase I, II and III transcription. The peptides act by binding to a
 CC different TAF, an activator, or TBP (TATA-binding protein) or
 CC competitively inhibiting association of a TAF domain with another
 CC compound, typically a protein like TBP or another TAF, an activator, or
 CC DNA. (Updated on 25-MAR-2003 to correct PF field.)
 XX
 SQ Sequence 1872 AA;

Query Match 48.5%; Score 48; DB 2; Length 1872;
 Best Local Similarity 62.5%; Pred. No. 1.9e+02;
 Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 AVRLLALSGSDGRVGS 19
 DT :::: |||| ||||
 DB 1740 AIQLSESGSDSDVGS 1755

RESULT 8
 AAR56491
 ID AAR56491 standard; protein; 1893 AA.
 XX
 AC AAR56491;

XX
 DT 25-MAR-2003 (revised)
 DT 23-MAR-1995 (first entry)
 XX
 DE TATA-binding protein-associated factor hTAFII250.
 XX
 KW TATA-binding protein associated factor; hTAFII250; screening; diagnostic;
 KW therapeutic; gene transcription regulation.
 XX
 OS Homo sapiens.
 XX
 PN WO9417087-A1.
 XX
 PD 04-AUG-1994.
 XX
 PF 28-JAN-1994; 94WO-US001114.
 XX
 PR 28-JAN-1993; 93US-00013412.
 PR 30-JUN-1993; 93US-00087119.
 XX
 PA (REGC) UNIV CALIFORNIA.
 XX
 PI Tjian R, Comai L, Dynlact BD, Hoey T, Ruppert S, Tanese N;
 PI Wang E, Weinzierl ROJ;
 PI
 DR WPI; 1994-264019/32.
 DR N-PSDB; AAQ70729.
 XX
 PT TATA-binding protein associated protein factors - and corresponding
 PT nucleotide sequence and deriv. antibodies, useful in screening,
 PT diagnostics and therapeutics.
 XX
 PS Disclosure; Page 112-123; 180pp; English.
 XX
 CC The TATA-binding protein associated factor hTAFII250 (including specific
 CC antibodies and fusion products) are used in drug screening, diagnostics
 CC and therapeutics. They are used in the development of specific
 CC biochemical assays for screening compounds that agonise or antagonise
 CC selected transcription factors involved in regulating gene expression
 CC associated with human pathology. (Updated on 25-MAR-2003 to correct PN
 CC field.)
 XX
 SQ Sequence 1893 AA;

Query Match 48.5%; Score 48; DB 2; Length 1893;
 Best Local Similarity 62.5%; Pred. No. 1.9e+02;
 Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 AVRLLALSGSDGRVGS 19
 DT :::: |||| ||||
 DB 1761 AIQLSESGSDSDVGS 1776

RESULT 9
 AAW06082
 ID AAW06082 standard; protein; 1893 AA.
 XX
 AC AAW06082;
 XX
 DT 25-MAR-2003 (revised)
 DT 27-JAN-1997 (first entry)
 XX
 DE Human TATA-binding protein associated factor hTAFII250 protein.
 XX
 KW Human; TATA-binding protein; TBP associated factor; TFIIID; holoenzyme;
 KW RNA polymerase II; transcription; messenger RNA; nuclear fraction;
 KW lambda-gt11; expression library.
 XX
 OS Homo sapiens.
 XX
 PN US5534410-A.
 XX
 PD 09-JUL-1996.

PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.

PS Claim 20; SEQ ID NO 36367; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activities. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 1924 AA;

Query Match 48.5%; Score 48; DB 4; Length 1924;
 Best Local Similarity 62.5%; Pred. No. 1.9e+02;
 Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 AVRLALSGSDGVGSG 19

Db 1792 AIQLSESGSDVDVSG 1807

RESULT 12

AAU43462
 ID AAU43462 standard; protein; 67 AA.

XX AAU43462;

DT 27-FEB-2002 (first entry)

XX Propionibacterium acnes immunogenic protein #4358.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KW dermatological; osteopathic; neuroprotectant.

XX Propionibacterium acnes.

PN WO200181581-A2.

PD 01-NOV-2001.

XX 20-APR-2001; 2001WO-US012865.

XX 21-APR-2000; 2000US-0199047P.

PR 02-JUN-2000; 2000US-0208841P.

PR 07-JUL-2000; 2000US-0216747P.

XX (CORI-) CORIXA CORP.

XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

PI L'Maisonneuve J, Zhang Y, Jen S, Carter D;

XX WPI; 2001-616774/71.

DR N-PSDB; AAS95920.

XX

PT Propionibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris.

PS Example 1; SEQ ID NO 4657; 1069pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
 CC polypeptides. The proteins and their associated DNA sequences are used in
 CC the treatment, prevention and diagnosis of medical conditions caused by
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
 CC P. acnes is also involved in infections of bone, joints and the central
 CC nervous system, however it is particularly involved in the inflammatory
 CC lesions associated with acne vulgaris. A method for detecting the
 CC presence or absence of P. acnes in a patient comprises contacting a
 CC sample with a binding agent that binds to the proteins of the invention
 CC and determining the amount of bound protein in the sample. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC specific for P. acnes proteins. These antibodies can be used to
 CC downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
 CC this patent did not form part of the printed specification, but was
 CC obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX Sequence 67 AA;

Query Match 47.5%; Score 47; DB 4; Length 67;
 Best Local Similarity 72.7%; Pred. No. 6.5;
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 10 SGSDGRVSGGC 20

Db 31 AGSEGRVGLGC 41

RESULT 13

ABM39981
 ID ABM39981 standard; protein; 67 AA.

XX ABM39981;

DT 20-OCT-2003 (first entry)

XX Propionibacterium acnes predicted ORF-encoded polypeptide #4657.

XX Acne vulgaris; antiseborrheic; dermatological; antibacterial;
 KW immunostimulant; immune response; vaccine.

XX Propionibacterium acnes.

PN WO2003033515-A1.

PD 24-APR-2003.

XX 11-OCT-2002; 2002WO-US032727.

PR 15-OCT-2001; 2001US-00978825.

XX (CORI-) CORIXA CORP.

XX Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;
 PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;

PI Barth B, Vallieue-Douglas J;

XX WPI; 2003-381789/36.

DR N-PSDB; ACP64449.

PT New Propionibacterium acnes polypeptides and polynucleotides encoding the
 PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,

PT or for stimulating an immune response specific for a P. acnes protein.
 XX
 PS Example 1; SEQ ID NO 4657; 1481pp; English.
 XX
 CC The invention relates to an isolated polynucleotide (ACF64435-ACF64733) encoding a Propionibacterium acnes protein. The invention also relates to
 CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
 CC immunogenic fragments of P. acnes polypeptides. The invention
 CC additionally encompasses expression vectors and host cells comprising a
 CC polynucleotide of the invention; antibodies against polypeptides of the
 CC invention; fusion proteins comprising a polypeptide of the invention; a
 CC method for stimulating an immune response specific for a P. acnes
 CC polypeptide and an isolated T cell population comprising T cells prepared
 CC via this method; a vaccine composition (comprising P. acnes polypeptides,
 CC polynucleotides, antibodies, fusion proteins, T cell populations, or
 CC antigen-presenting cells that express the polypeptide); a method and kit
 CC for detecting or determining the presence or absence of P. acnes in a
 CC patient; and a method for inhibiting the development of P. acnes in a
 CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
 CC proteins, T cell populations or antigen-presenting cells that express the
 CC polypeptides are useful for diagnosing, preventing or treating acne
 CC vulgaris, or for stimulating an immune response specific for a P. acnes
 CC protein. The polynucleotides can also be used as probes or primers for
 CC nucleic acid hybridisation. The vaccine composition is useful for the
 CC stimulation of an immune response against P. acnes, or for treating acne,
 CC and the kit is useful for performing a diagnostic assay. The present
 CC sequence represents a polypeptide predicted to be encoded by an ORF (open
 CC reading frame) contained within the P. acnes polynucleotides of the
 CC invention. Note: The sequence data for this patent did not form part of
 CC the printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 67 AA;

Query Match 47.5%; Score 47; DB 6; Length 67;
 Best Local Similarity 72.7%; Pred. No. 6.5;
 Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 10 SCSDGRVSGC 20
 :||:|||||
 Db 31 AGSEGRVGLGC 41

RESULT 14
 ABG29258
 ID ABG29258 standard; protein; 220 AA.

XX ABG29258;
 XX
 XX 18-FEB-2002 (first entry)
 DT
 XX Novel human diagnostic protein #29249.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

OS WO200175067-A2.

PN 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US0008631.

XX 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

DR N-PSDB; AAS93445.

XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX

PS Claim 20; SEQ ID NO 59617; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
 CC amino acid sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 220 AA;

Query Match 46.5%; Score 46; DB 4; Length 220;
 Best Local Similarity 61.1%; Pred. No. 36;
 Matches 11; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 2 ORAVRLALSGSDGRVSG 19
 |||||
 Db 195 QRAVRKALPLSGSGTGAG 212

RESULT 15

ABB48949

ID ABB48949 standard; protein; 352 AA.

XX ABB48949;

XX 05-FEB-2002 (first entry)

XX Listeria monocytogenes protein #1653.

XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
 KW vitamin B12; bacterial infection; disease.

XX Listeria monocytogenes.

OS WO200177335-A2.

PN 18-OCT-2001.

XX 11-APR-2001; 2001WO-FR001118.

XX 11-APR-2000; 2000FR-00004629.

XX (INSP) INST PASTEUR.

XX Buchrieser C, Frangeul L, Couve E, Rueniok C, Feihl H, Dehoux P;
 PI Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst P, Cossart P;
 PI Daniels J, Goebel W, Krefit J, Kuhn M, Ng E, Vazquez-Boland JA;
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
 PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
 PI Maduenio E, De Pablo B, Wehland J, Kaerst U, Entian K, Hauf J;

PI Rose M, Voss H;
XX WPI; 2002-010914/01.
XX Genomic sequence for *Listeria monocytogenes*, useful e.g. for treatment
PT and prevention of *Listeria* and related bacterial infections, and related
PT polypeptides.
XX Claim 6; SEQ ID NO 1654; 192pp; French.
XX The present invention relates to the genome sequence of *Listeria*
CC *monocytogenes* EGD-e (see ABA03041). The genome sequence and fragments of
CC it are useful for selecting probes and primers for detecting genes in *L.*
CC *monocytogenes* and related organisms, and for studying genetic
CC polymorphisms and other genomes. The present sequence is a protein
CC encoded by the genome sequence of the present invention. Proteins
CC expressed from the genome sequence are useful for raising specific
CC antibodies, identification of *L. monocytogenes* and related organisms, and
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
CC B12. The genome sequence and proteins encoded by it are also useful for
CC selecting compounds that regulate gene expression and cell replication
CC and modulate *L. monocytogenes*-related diseases. In addition, the genome
CC sequence and proteins encoded by it are useful in pharmaceutical and
CC vaccines compositions for the treatment or prevention of infections by *L.*
CC *monocytogenes* and related organisms. Note: the sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 352 AA;

Query Match 46.5%; Score 46; DB 5; Length 352;
Best Local Similarity 71.4%; Pred. No. 60;
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 RAVRLALSGDGRV 16
DB 119 RAVLLALGGADGHV 132

RESULT 16
AAM00762
ID AAM00762 standard; peptide; 17 AA.
XX
AC AAM00762;
XX
DT 01-OCT-2001 (first entry)
XX
DE Human bone marrow protein, SEQ ID NO: 125.
XX
KW Human; bone marrow; antiinflammatory; cytostatic; neuroprotective;
KW antiviral; antibacterial; antifungal; anti-HIV; haemostatic;
KW immunosuppressive; gene therapy; cytokine cell proliferation;
KW cell differentiation modulator; immune disorder; infection; cancer;
KW human immunodeficiency virus; HIV; autoimmune disorder; haemophilia.
XX
OS Homo sapiens.
XX
PN WO200153453-A2.
XX
PD 26-JUL-2001.
XX
XX 23-DEC-2000; 2000WO-US034960.
XX
PR 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-0048725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 30-NOV-2000; 2000US-0250583P.

XX (HYSE-) HYSEQ INC.
XX Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Chen R, Ma Y;
PI Ren F, Wang J, Werhman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Drmanac RT;
XX WPI; 2001-488707/53.
DR N-PSDB; AAH89881.
XX
XX Novel bone marrow-expressed polynucleotides and polypeptides, useful for
PT treating e.g. cancer and immune deficiency disorders.
XX Claim 10; Page 296; 648pp; English.
XX The present sequence is one of 251 novel human polypeptides encoded by a
CC bone marrow-expressed polynucleotide. The polynucleotide and the
CC polypeptide encoded by it are useful in the treatment of various immune
CC deficiencies and disorders. The deficiencies and disorders may be
CC genetic, may be caused by a viral (e.g. HIV), bacterial or fungal
CC infection, or may result from an autoimmune disorder, a coagulation
CC disorder (e.g. haemophilia), inhibition of tumour cell proliferation,
CC suppression of an inflammatory response or treatment of a nervous system
CC disorder such as Alzheimer's disease. Detection of the presence or
CC increased expression of the polynucleotide or the protein it encodes is
CC useful for the diagnosis and/or prognosis of one or more types of cancer.
CC The polynucleotide and polypeptide can be used as nutritional sources or
CC supplements and in the screening of chemical compounds as potential drugs
XX
SQ Sequence 17 AA;

Query Match 45.5%; Score 45; DB 4; Length 17;
Best Local Similarity 57.1%; Pred. No. 2.9;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 6 RLALSGDGRVCSG 19
DB 3 RWAMAGEDGKVGCG 16

RESULT 17
AAU33620
ID AAU33620 standard; protein; 359 AA.
XX
AC AAU33620;
XX
DT 14-FEB-2002 (first entry)
XX
DE *Pseudomonas aeruginosa* cellular proliferation protein #64.
XX
KW Antisense; prokaryotic cellular proliferation protein; antibiotic;
KW antibacterial; drug design.
XX
OS *Pseudomonas aeruginosa*.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
XX 21-MAR-2001; 2001WO-US009180.
XX
PR 21-MAR-2000; 2000US-0191078P.
PR 23-MAY-2000; 2000US-0206848P.
PR 26-MAY-2000; 2000US-0207727P.
PR 23-OCT-2000; 2000US-0242578P.
PR 27-NOV-2000; 2000US-0253625P.
PR 22-DEC-2000; 2000US-0257931P.
PR 16-FEB-2001; 2001US-0269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;

```

XX WPI; 2001-611495/70.
DR N-PSDB; AAS51479.
XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids.
PT
XX Example 3; SEQ ID NO 5116; 511pp; English.
PS
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the genes,
CC their use in the discovery of novel antibiotics, the essential genes
CC themselves and the encoded proteins. The prokaryotes used are Escherichia
CC coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
CC Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
CC useful for the identification of potential new targets for antibiotic
CC development. The antisense nucleic acids can also be used to identify
CC proteins used in proliferation, to express these proteins, and to obtain
CC antibodies capable of binding to the expressed proteins. The proteins can
CC be used to screen compounds in rational drug discovery programmes. The
CC antisense nucleic acid sequence is also useful to screen for homologous
CC nucleic acids which are required for cell proliferation in a wide variety
CC of organisms. The present sequence represents an essential prokaryotic
CC cellular proliferation protein. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 359 AA;
Query Match 45.5%; Score 45; DB 4; Length 359;
Best Local Similarity 62.5%; Pred. No. 90;
Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 1 LQRAVLALSGSDGRV 16
DB 126 LDRSARLSLGGIDPRV 141
RESULT 18
ABU15565
ID ABU15565 standard; protein; 359 AA.
XX
AC ABU15565;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #1092.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Pseudomonas aeruginosa.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
XX WPI; 2003-029926/02.
DR N-PSDB; ACA19435.
XX

```

```

PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 43489; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: the sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 359 AA;
Query Match 45.5%; Score 45; DB 6; Length 359;
Best Local Similarity 62.5%; Pred. No. 90;
Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 1 LQRAVLALSGSDGRV 16
DB 126 LDRSARLSLGGIDPRV 141
RESULT 19
ABBS59456
ID ABBS59456 standard; protein; 394 AA.
XX
AC ABBS59456;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 5160.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
PD
XX 23-MAR-2001; 2001WO-US009231.
PF
XX 23-MAR-2000; 2000US-0191637P.
PR
XX 11-JUL-2000; 2000US-00614150.
XX

```

PA (PEKE) PE CORP NY.
 XX Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656860/75.
 DR N-PSDB; ABL03559.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 XX Disclosure; SEQ ID NO 5160; 21pp + Sequence Listing; English.
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX Sequence 394 AA;
 SQ Query Match 45.5%; Score 45; DB 4; Length 394;
 Best Local Similarity 40.0%; Pred. No. 1e+02;
 Matches 8; Conservative 6; Mismatches 6; Indels 0; Gaps 0;
 QY 1 LQRAVRLALSGSDGRVSGC 20
 DB 177 LAKSLKLVKTGTSFHVSGC 196
 RESULT 20
 AAY22514
 ID AAY22514 standard; protein; 73 AA.
 XX AC AAY22514;
 XX DT 19-OCT-1999 (first entry)
 XX Xenopus hnRNP C1/C2 protein.
 DE Cognate RNA; gene expression regulator; RNA-protein interaction; U1 RNA;
 KW autoimmune disease; neoplastic disease; developmental biology;
 KW RNA-associated protein.
 OS Xenopus sp.
 XX US5866680-A.
 XX PD 02-FEB-1999.
 XX PF 07-JUN-1995; 95US-00474753.
 XX 15-NOV-1989; 89US-00436779.
 PR 12-JUN-1990; 90US-00535943.
 PR 27-DEC-1993; 93US-00173941.
 XX (KEEN/) KEENE J D.
 PA (QUER/) QUERY C C.
 PA (BENT/) BENTLEY R.
 XX Bentley RO, Keene JD, Query CC;
 PI WPI; 1999-142012/12.
 DR Conjugates of U1 ribonucleoprotein RNA - with nucleic acid probes,
 PT antigens and antibodies.
 XX Claim 6; Fig 6; 53pp; English.

XX This sequence represents an RNA-associated protein. The invention relates
 CC to a compound of formula (IX): (component Y) - (cognate RNA) where:
 CC component Y is covalently bound to the cognate RNA and is selected from
 CC DNA probes, RNA probes, antigens and antibodies; and the cognate RNA is a
 CC polynucleotide comprising at least the sequence shown in AAX84881 and up
 CC to the sequence of U1 RNA. The compounds may have a useful application in
 CC numerous fields, including: regulation of gene expression, RNA-protein
 CC interactions, autoimmune and neoplastic diseases, and developmental
 XX biology
 SQ Sequence 73 AA;
 Query Match 44.4%; Score 44; DB 2; Length 73;
 Best Local Similarity 44.4%; Pred. No. 22;
 Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;
 QY 2 QRAVRLALSGSDGRVSG 19
 DB 44 ERTARTAVAGEDGRMIAG 61
 RESULT 21
 ABB63810
 ID ABB63810 standard; protein; 158 AA.
 XX AC ABB63810;
 XX DT 26-MAR-2002 (first entry)
 XX Drosophila melanogaster polypeptide SEQ ID NO 18222.
 DE Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 OS Drosophila melanogaster.
 XX WO200171042-A2.
 XX PD 27-SEP-2001.
 XX PF 23-MAR-2001; 2001WO-US009231.
 XX 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX (PEKE) PE CORP NY.
 XX Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656860/75.
 DR N-PSDB; ABL07913.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 XX Disclosure; SEQ ID NO 18222; 21pp + Sequence Listing; English.
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX Sequence 158 AA;
 SQ Query Match 44.4%; Score 44; DB 4; Length 158;

Best Local Similarity 58.8%; Pred. No. 52;
Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 4 AVRLALSGSDGRVSGC 20
DB 47 AVGAALRTSDGTIYSGC 63

RESULT 22
ADC07742
ID ADC07742 standard; protein; 133 AA.

AC ADC07742;
DT 18-DEC-2003 (first entry)

DE Rice protein sequence Seq ID8 related to grain filling.

XX plant biotechnology; carbohydrate synthesis; carbohydrate metabolism;
KW carbohydrate degradation; carbohydrate; plant grain; grain filling; corn;
KW tomato; banana; canola; cotton; peanut; sorghum; tobacco; sugarbeet;
KW wheat; rice; protein; oil; starch; fibre; moisture content; cereal grain;
KW Gene; ds; plant.

OS Oryza sativa.

XX WO2003000905-A2.

PN 03-JAN-2003.

XX 21-JUN-2002; 2002WO-IB002450.

XX 22-JUN-2001; 2001US-0300112P.

PR 26-SEP-2001; 2001US-0325277P.

PR 20-DEC-2001; 2001US-0342327P.

XX (SYGN) SYNGENTA PARTICIPATIONS AG.

XX Zhu T, Cheng W, Briggs S, Cooper B, Goff SA, Moughamer T;
PI Glazebrook J, Katagiri F, Provart N, Ricke D;

XX WPI; 2003-229341/22.

DR N-PSDB; ADC07741.

XX New plant genes encoding polypeptides having an activity involved in or
PT associated with the synthesis, metabolism or degradation of carbohydrates
PT in the plant grain useful in generating plants having improved
PT nutritional properties.

PS Claim 34; SEQ ID NO 8; 130pp; English.

XX This invention, in the area of plant biotechnology, relates to novel
CC polynucleotides comprising a nucleotide sequence encoding a protein which
CC is involved in or associated with the synthesis, metabolism or
CC degradation of carbohydrates in the plant grain and the expression of
CC which is up-regulated during grain filling. The plant is selected from
CC corn, tomato, banana, canola, cotton, peanut, sorghum, tobacco,
CC sugarbeet, wheat, and rice. The invention may be useful for the
CC improvement of protein, oil, starch, fibre and moisture content of the
CC cereal grains. In addition, carbohydrate levels may be modified to a more
CC desirable level using the present invention. The present sequence is the
CC amino acid sequence of a rice protein of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/publishedseq_sequences.

XX Sequence 133 AA;

Query Match 43.4%; Score 43; DB 7; Length 133;
Best Local Similarity 61.5%; Pred. No. 62;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 7 LALSGSDGRVSG 19

DB 18 LAATGXDGRAGDG 30
RESULT 23
AAG46917
ID AAG46917 standard; protein; 215 AA.

XX AAG46917;

XX 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 59076.

XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX Arabidopsis thaliana.

PN EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-00301439.

XX 25-FEB-1999; 99US-0121825P.

PR 05-MAR-1999; 99US-0123180P.

PR 09-MAR-1999; 99US-0123548P.

PR 23-MAR-1999; 99US-0125788P.

PR 25-MAR-1999; 99US-0126264P.

PR 29-MAR-1999; 99US-0126785P.

PR 01-APR-1999; 99US-0127462P.

PR 08-APR-1999; 99US-0128234P.

PR 16-APR-1999; 99US-0128714P.

PR 19-APR-1999; 99US-0130077P.

PR 21-APR-1999; 99US-0130449P.

PR 23-APR-1999; 99US-0130510P.

PR 28-APR-1999; 99US-0131449P.

PR 30-APR-1999; 99US-0132048P.

PR 04-MAY-1999; 99US-0132407P.

PR 05-MAY-1999; 99US-0132484P.

PR 06-MAY-1999; 99US-0132485P.

PR 06-MAY-1999; 99US-0132486P.

PR 11-MAY-1999; 99US-0132863P.

PR 14-MAY-1999; 99US-0134218P.

PR 14-MAY-1999; 99US-0134219P.

PR 14-MAY-1999; 99US-0134221P.

PR 18-MAY-1999; 99US-0134370P.

PR 19-MAY-1999; 99US-0134768P.

PR 20-MAY-1999; 99US-0134941P.

PR 21-MAY-1999; 99US-0135124P.

PR 21-MAY-1999; 99US-0135353P.

PR 25-MAY-1999; 99US-0136021P.

PR 27-MAY-1999; 99US-0136392P.

PR 28-MAY-1999; 99US-0136782P.

PR 01-JUN-1999; 99US-0137222P.

PR 03-JUN-1999; 99US-0137528P.

PR 04-JUN-1999; 99US-0137502P.

PR 07-JUN-1999; 99US-0137724P.

PR 08-JUN-1999; 99US-0138094P.

PR 10-JUN-1999; 99US-0138540P.

PR 14-JUN-1999; 99US-0138847P.

PR 16-JUN-1999; 99US-0139119P.

PR 16-JUN-1999; 99US-0139452P.

PR 17-JUN-1999; 99US-0139453P.

PR 18-JUN-1999; 99US-0139492P.

PR 18-JUN-1999; 99US-0139454P.

PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 24-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142820P.
PR 12-JUL-1999; 99US-0142377P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 21-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145192P.
PR 22-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145219P.
PR 26-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 28-JUL-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 04-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149368P.

PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 43.4%; Score 43; DB 3; Length 215;
Best Local Similarity 38.9%; Pred. No. 1.1e+02;
Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 2 QRAVRLALSGSGRGVSG 19
Db 71 RKSIRVTQGTGRSGS 88

RESULT 24
ABM67849
ID ABM67849 standard; protein; 233 AA.

XX AC ABM67849;
XX DT 20-NOV-2003 (first entry)
XX DE
XX DE Photorhabdus luminescens protein sequence #946.
XX KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW detection; food; gene expression; plant; animal; microorganism; toxin;
KW antibiotic; biopesticide; virulence factor; disease model; plague;
KW whooping cough.
XX OS
XX OS Photorhabdus luminescens.
XX PN WO200294867-A2.
XX PD 28-NOV-2002.
XX PF 07-FEB-2002; 2002WO-IB003040.
XX PR 07-FEB-2001; 2001FR-00001659.
XX PA (INSP) INST PASTEUR.
XX PA (CNRS) CNRS CENT NAT RECH SCI.
XX PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst P, Danchin A,
PI Buchrieser C;
XX XX
XX DR WPI; 2003-148459/14.
XX PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
XX useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX PS Claim 2; SEQ ID NO 946; 1205pp; French.
XX CC The invention relates to the isolation of genes and their encoded
CC proteins from Photorhabdus luminescens. The isolated sequences are
CC sources of probes and primers for detecting the genome of P. luminescens
CC and related species; to study polymorphisms; for gene analysis and for
CC detection/amplification of the genes. Antibodies (Ab) raised against the
CC polypeptides encoded by the genes are used for detection/identification
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC carry a gene-containing vector are used to select compounds that
CC modulate, regulate, induce or inhibit expression of the genes in plants,
CC animals or microorganisms other than P. luminescens and are able to alter
CC response or sensitivity to toxins and antibiotics produced by P.
CC luminescens. Cells transformed to express the genes are useful for
CC recombinant production of the proteins, particularly toxins and
CC antibacterials useful as insecticides, bactericides and fungicides. The
CC genes, proteins, vectors containing the genes and Ab are also useful
CC therapeutically (to treat microbial infection by bacteria or fungi that
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC biopesticides. Other uses of the genes and the proteins are as virulence
CC factors and for identifying targets of human diseases for which P.
CC luminescens is a model (particularly plague and whooping cough). This
CC sequence represents one of the isolated P. luminescens proteins
XX SQ Sequence 233 AA;
Query Match 43.4%; Score 43; DB 6; Length 233;
Best Local Similarity 64.3%; Pred. No. 1.2e+02;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 3 RAVRLALSGSDGRV 16
||:|||||:
Db 110 RAIRLALEGVDRRL 123
RESULT 25
AAG46916
ID AAG46916 standard; protein; 258 AA.
XX AAG46916;
AC AAG46916;

XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 59075.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-00301439.
XX PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 29-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130449P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 06-MAY-1999; 99US-0132487P.
PR 07-MAY-1999; 99US-0132863P.
PR 11-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135533P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 10-JUN-1999; 99US-0138847P.
PR 14-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 16-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139452P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.

PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140855P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-01411287P.
PR 01-JUL-1999; 99US-0141842P.
PR 02-JUL-1999; 99US-0142154P.
PR 06-JUL-1999; 99US-0142055P.
PR 08-JUL-1999; 99US-0142390P.
PR 09-JUL-1999; 99US-0142803P.
PR 12-JUL-1999; 99US-0142920P.
PR 13-JUL-1999; 99US-0142977P.
PR 14-JUL-1999; 99US-0143542P.
PR 15-JUL-1999; 99US-0143624P.
PR 16-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 19-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 02-AUG-1999; 99US-0146389P.
PR 03-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 05-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 05-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 06-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 09-AUG-1999; 99US-0147935P.
PR 10-AUG-1999; 99US-0148171P.
PR 11-AUG-1999; 99US-0148319P.
PR 12-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 13-AUG-1999; 99US-0148684P.
PR 16-AUG-1999; 99US-0149168P.
PR 17-AUG-1999; 99US-0149175P.
PR 18-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.

PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 22-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 23-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 43.4%; Score 43; DB 3; Length 258;
Best Local Similarity 38.9%; Pred. No. 1.3e+02;

Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 2 GRAVRLALSGSGRGVSG 19
Db 114 RKSIRVTVGRTGRSGG 131
:::|:|:|

RESULT 26

AAG46915
ID AAG46915 standard; protein; 297 AA.

XX AAG46915;

AC AAG46915;

DT 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 59074.

XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;

termination sequence.
Arabidopsis thaliana.
EP1033405-A2.
06-SEP-2000.
25-FEB-2000; 2000EP-00301439.
25-FEB-1999; 99US-0121825P.
05-MAR-1999; 99US-0123180P.
09-MAR-1999; 99US-0123548P.
23-MAR-1999; 99US-0125788P.
25-MAR-1999; 99US-0126264P.
29-MAR-1999; 99US-0126785P.
01-APR-1999; 99US-0127462P.
06-APR-1999; 99US-0128234P.
08-APR-1999; 99US-0128714P.
16-APR-1999; 99US-0129845P.
19-APR-1999; 99US-0130077P.
21-APR-1999; 99US-0130449P.
23-APR-1999; 99US-0130510P.
23-APR-1999; 99US-0130891P.
28-APR-1999; 99US-0131449P.
30-APR-1999; 99US-0132048P.
30-APR-1999; 99US-0132407P.
04-MAY-1999; 99US-0132484P.
05-MAY-1999; 99US-0132485P.
06-MAY-1999; 99US-0132486P.
07-MAY-1999; 99US-0132487P.
11-MAY-1999; 99US-0132863P.
11-MAY-1999; 99US-0134256P.
14-MAY-1999; 99US-0134218P.
14-MAY-1999; 99US-0134219P.
14-MAY-1999; 99US-0134221P.
14-MAY-1999; 99US-0134370P.
18-MAY-1999; 99US-0134768P.
19-MAY-1999; 99US-0134941P.
20-MAY-1999; 99US-0135124P.
21-MAY-1999; 99US-0135353P.
24-MAY-1999; 99US-0135629P.
25-MAY-1999; 99US-0136021P.
27-MAY-1999; 99US-0136392P.
28-MAY-1999; 99US-0136782P.
01-JUN-1999; 99US-0137222P.
03-JUN-1999; 99US-0137528P.
04-JUN-1999; 99US-0137502P.
07-JUN-1999; 99US-0137724P.
08-JUN-1999; 99US-0138094P.
10-JUN-1999; 99US-0138540P.
10-JUN-1999; 99US-0138847P.
14-JUN-1999; 99US-0139119P.
16-JUN-1999; 99US-0139452P.
16-JUN-1999; 99US-0139453P.
17-JUN-1999; 99US-0139492P.
18-JUN-1999; 99US-0139454P.
18-JUN-1999; 99US-0139455P.
18-JUN-1999; 99US-0139456P.
18-JUN-1999; 99US-0139457P.
18-JUN-1999; 99US-0139458P.
18-JUN-1999; 99US-0139459P.
18-JUN-1999; 99US-0139460P.
18-JUN-1999; 99US-0139461P.
18-JUN-1999; 99US-0139462P.
18-JUN-1999; 99US-0139463P.
18-JUN-1999; 99US-0139750P.
18-JUN-1999; 99US-0139763P.
21-JUN-1999; 99US-0139817P.
22-JUN-1999; 99US-0139899P.
23-JUN-1999; 99US-0140353P.
23-JUN-1999; 99US-0140354P.
24-JUN-1999; 99US-0140695P.
28-JUN-1999; 99US-0140823P.
29-JUN-1999; 99US-0140991P.
30-JUN-1999; 99US-0141287P.
01-JUL-1999; 99US-0141842P.
01-JUL-1999; 99US-0142154P.
02-JUL-1999; 99US-0142055P.
06-JUL-1999; 99US-0142390P.
08-JUL-1999; 99US-0142803P.
09-JUL-1999; 99US-0142920P.
12-JUL-1999; 99US-0142977P.
13-JUL-1999; 99US-0143542P.
14-JUL-1999; 99US-0143624P.
15-JUL-1999; 99US-0144005P.
16-JUL-1999; 99US-0144085P.
16-JUL-1999; 99US-0144086P.
19-JUL-1999; 99US-0144325P.
19-JUL-1999; 99US-0144331P.
19-JUL-1999; 99US-0144332P.
19-JUL-1999; 99US-0144333P.
19-JUL-1999; 99US-0144334P.
19-JUL-1999; 99US-0144335P.
20-JUL-1999; 99US-0144352P.
20-JUL-1999; 99US-0144632P.
20-JUL-1999; 99US-0145087P.
21-JUL-1999; 99US-0144884P.
21-JUL-1999; 99US-0144814P.
21-JUL-1999; 99US-0145086P.
21-JUL-1999; 99US-0145088P.
22-JUL-1999; 99US-0145085P.
22-JUL-1999; 99US-0145087P.
22-JUL-1999; 99US-0145089P.
22-JUL-1999; 99US-0145192P.
23-JUL-1999; 99US-0145145P.
23-JUL-1999; 99US-0145218P.
23-JUL-1999; 99US-0145224P.
26-JUL-1999; 99US-0145276P.
27-JUL-1999; 99US-0145913P.
27-JUL-1999; 99US-0145918P.
27-JUL-1999; 99US-0145919P.
28-JUL-1999; 99US-0145951P.
02-AUG-1999; 99US-0146386P.
02-AUG-1999; 99US-0146388P.
02-AUG-1999; 99US-0146389P.
03-AUG-1999; 99US-0147038P.
04-AUG-1999; 99US-0147204P.
04-AUG-1999; 99US-0147302P.
05-AUG-1999; 99US-0147192P.
05-AUG-1999; 99US-0147260P.
06-AUG-1999; 99US-0147303P.
06-AUG-1999; 99US-0147416P.
09-AUG-1999; 99US-0147493P.
09-AUG-1999; 99US-0147935P.
10-AUG-1999; 99US-0148171P.
11-AUG-1999; 99US-0148319P.
12-AUG-1999; 99US-0148341P.
13-AUG-1999; 99US-0148565P.
13-AUG-1999; 99US-0148684P.
16-AUG-1999; 99US-0149368P.
17-AUG-1999; 99US-0149175P.
18-AUG-1999; 99US-0149426P.
20-AUG-1999; 99US-0149722P.
20-AUG-1999; 99US-0149723P.
20-AUG-1999; 99US-0149929P.
23-AUG-1999; 99US-0149902P.
23-AUG-1999; 99US-0149930P.
25-AUG-1999; 99US-0150566P.
26-AUG-1999; 99US-0150884P.
27-AUG-1999; 99US-0151065P.
27-AUG-1999; 99US-0151066P.
27-AUG-1999; 99US-0151080P.
30-AUG-1999; 99US-0151303P.
31-AUG-1999; 99US-0151438P.
01-SEP-1999; 99US-0151930P.
07-SEP-1999; 99US-0152363P.

```

PR 10-SEP-1999; 99US-0153070P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159638P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 22-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160880P.
PR 22-OCT-1999; 99US-0160981P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 43.4%; Score 43; DB 3; Length 297;
Best Local Similarity 38.9%; Pred. No. 1.5e+02;
Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 2 QRAVRLALSGSDGRVGS 19
Db 153 RKSIRVTQGTGRSGS 170

RESULT 27
ABG11319
ID ABG11319 standard; protein; 349 AA.
XX
AC ABG11319;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #11310.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX

```

```

PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS75506.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 20; SEQ ID NO 41678; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological actions. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 349 AA;

Query Match 43.4%; Score 43; DB 4; Length 349;
Best Local Similarity 42.1%; Pred. No. 1.8e+02;
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 1 LORAVRLALSGSDGRVGS 19
Db 94 MKRNARAAVAGEDGRMIAG 112

RESULT 28
ABM68505
ID ABM68505 standard; protein; 436 AA.
XX
AC ABM68505;
XX
DT 20-NOV-2003 (first entry)
XX
DE Photorhabdus luminescens protein sequence #1602.
XX
KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW detection; food; gene expression; plant; animal; microorganism; toxin;
KW antibiotic; biopesticide; virulence factor; disease model; plague;
XX
OS Photorhabdus luminescens.
XX
PN WO200294867-A2.
XX
PD 28-NOV-2002.
XX

```


CC AAB99143, AAH75903-AAH75920 and AAG66436

XX
SQ Sequence 1017 AA;

Query Match 43.4%; Score 43; DB 4; Length 1017;
Best Local Similarity 58.8%; Pred. No. 6.1e+02;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 QRAVRLALSGDGRVGS 18

Db 798 ERAGRVVTLGSDYRVGS 814

Search completed: May 13, 2004, 06:52:39
Job time : 52 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2004, 06:49:48 ; Search time 14 Seconds
(without alignments)
73.751 Million cell updates/sec

Title: 09549186-9

Perfect score: 99

Sequence: 1 LQRAVRLALSGDGRVSGSC 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Issued Patents AA:*

1: /cgn2.6/prodata/2/iaa/5A_COMB.pep:*

2: /cgn2.6/prodata/2/iaa/5B_COMB.pep:*

3: /cgn2.6/prodata/2/iaa/6A_COMB.pep:*

4: /cgn2.6/prodata/2/iaa/6B_COMB.pep:*

5: /cgn2.6/prodata/2/iaa/PTUS_COMB.pep:*

6: /cgn2.6/prodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	68	68.7	19	2	US-08-374-560-7
2	48	48.5	1872	1	US-08-188-582-14
3	48	48.5	1872	1	US-08-646-715-14
4	48	48.5	1893	1	US-08-188-582-11
5	48	48.5	1893	1	US-08-646-715-11
6	45.5	46.0	341	4	US-09-252-991A-30646
7	45	45.5	212	4	US-09-252-991A-27887
8	45	45.5	432	4	US-09-252-991A-31177
9	44.5	44.9	660	4	US-09-252-991A-32206
10	44	44.4	546	4	US-09-252-991A-1901
11	43.5	43.9	410	4	US-09-252-991A-1891
12	42.5	42.9	802	4	US-09-252-991A-25050
13	42	42.4	134	4	US-09-252-991A-23103
14	42	42.4	187	4	US-09-673-395A-235
15	42	42.4	195	4	US-09-673-395A-571
16	42	42.4	224	4	US-09-673-395A-324
17	42	42.4	474	4	US-09-252-991A-19401
18	42	42.4	547	3	US-09-176-657-1
19	42	42.4	547	4	US-09-421-299-1
20	42	42.4	548	4	US-09-205-258-264
21	42	42.4	1477	3	US-08-492-459-10
22	42	42.4	1477	3	US-08-423-752-10
23	42	42.4	1477	3	US-08-945-994-3
24	42	42.4	1477	3	US-08-716-873-24
25	42	42.4	1477	3	US-09-368-431-24
26	42	42.4	1477	4	US-09-414-006-10
27	42	42.4	1477	4	US-09-447-223-10

41.5	41.9	1495	4	US-09-543-681A-5986	Sequence 5986, Ap
41	41.4	240	1	US-08-362-670B-30	Sequence 30, Appl
30	41	240	3	US-08-333-576C-30	Sequence 30, Appl
31	41.4	240	3	US-08-808-324-30	Sequence 30, Appl
32	41	240	5	PCT-US94-14030A-30	Sequence 30, Appl
33	41	374	4	US-09-252-991A-29999	Sequence 29999, A
34	41.4	376	4	US-09-252-991A-31366	Sequence 31366, A
35	41	400	4	US-09-252-991A-32313	Sequence 32313, A
36	41	402	4	US-09-252-991A-20914	Sequence 20914, A
37	41	465	4	US-09-252-991A-18919	Sequence 18919, A
38	41	469	4	US-09-252-991A-25490	Sequence 25490, A
39	41	600	2	US-08-821-118-19	Sequence 19, Appl
40	41	600	2	US-09-029-047C-2	Sequence 2, Appl
41	41.4	2644	4	US-09-328-352-4147	Sequence 4147, Ap
42	40	105	4	US-09-252-991A-23248	Sequence 23248, A
43	40	160	4	US-09-252-991A-27862	Sequence 27862, A
44	40	184	4	US-09-252-991A-22597	Sequence 22597, A
45	40	216	4	US-09-252-991A-27818	Sequence 27818, A
46	40	241	4	US-09-252-991A-32319	Sequence 32319, A
47	40	272	5	PCT-US95-02455-4	Sequence 4, Appl
48	40	278	4	US-09-489-039A-8998	Sequence 8998, Ap
49	40	279	4	US-09-252-991A-23964	Sequence 23964, A
50	40	292	4	US-09-252-991A-29835	Sequence 29835, A
51	40	320	4	US-09-252-991A-21006	Sequence 21006, A
52	40	335	4	US-09-540-236-3638	Sequence 3638, Ap
53	40	338	4	US-09-252-991A-19609	Sequence 19609, A
54	40	343	4	US-09-328-352-4321	Sequence 4321, Ap
55	40	444	4	US-09-489-039A-14180	Sequence 14180, A
56	40	446	4	US-09-543-681A-4620	Sequence 4620, Ap
57	40	521	4	US-09-252-991A-25791	Sequence 25791, A
58	40	580	4	US-09-252-991A-20407	Sequence 20407, A
59	40	673	4	US-09-252-991A-28817	Sequence 28817, A
60	40	780	4	US-09-252-991A-18846	Sequence 18846, A
61	40	812	4	US-09-252-991A-30710	Sequence 30710, A
62	40	812	4	US-09-252-991A-20256	Sequence 20256, A
63	39.5	316	4	US-09-252-991A-16444	Sequence 16444, App
64	39.5	436	4	US-09-252-991A-18918	Sequence 18918, A
65	39.5	471	4	US-09-252-991A-25549	Sequence 25549, A
66	39.5	508	4	US-09-252-991A-18886	Sequence 18886, A
67	39.5	134	4	US-08-581-528A-6	Sequence 6, Appl
68	39	161	2	US-09-412-791D-6	Sequence 6, Appl
69	39	161	4	PCT-US94-07799-6	Sequence 6, Appl
70	39	161	4	US-09-071-035-284	Sequence 284, App
71	39	169	4	US-09-252-991A-21582	Sequence 21582, A
72	39	227	4	US-09-711-164-447	Sequence 447, App
73	39	229	4	US-09-489-039A-13515	Sequence 13515, A
74	39	237	4	US-09-071-035-282	Sequence 282, App
75	39	259	4	US-09-252-991A-17103	Sequence 17103, A
76	39	276	4	US-09-252-991A-22266	Sequence 22266, A
77	39	309	4	US-09-454-279-12	Sequence 12, Appl
78	39	331	4	US-09-328-352-7502	Sequence 7502, Ap
79	39	359	4	US-09-134-001C-3830	Sequence 3830, Ap
80	39	380	4	US-09-252-991A-18543	Sequence 18543, A
81	39	382	4	US-09-347-798-8	Sequence 8, Appl
82	39	391	3	US-09-697-367-22	Sequence 22, Appl
83	39	433	4	US-09-252-991A-18775	Sequence 18775, A
84	39	443	4	US-09-252-991A-28932	Sequence 28932, A
85	39	468	4	US-09-252-991A-25740	Sequence 25740, A
86	39	489	4	US-09-252-991A-21935	Sequence 21935, A
87	39	539	4	US-08-262-424-7	Sequence 7, Appl
88	39	568	1	US-08-493-197-7	Sequence 7, Appl
89	39	568	2	US-08-717-587A-4	Sequence 4, Appl
90	39	568	3	US-08-883-610A-4	Sequence 4, Appl
91	39	568	3	US-08-936-094A-4	Sequence 4, Appl
92	39	568	5	PCT-US95-07844-7	Sequence 7, Appl
93	39	709	4	US-09-328-352-5172	Sequence 5172, Ap
94	39	921	4	US-09-252-991A-32974	Sequence 32974, A
95	39	94	4	US-09-252-991A-27341	Sequence 27341, A
96	39	1073	4	US-09-410-551B-21	Sequence 21, Appl
97	39	1482	4	US-09-410-551B-17	Sequence 17, Appl
98	39	1488	4	US-09-410-551B-23	Sequence 23, Appl
99	39	1509	4		
100	39				


```

; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1893 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-188-582-11

Query Match          48.5%; Score 48; DB 1; Length 1893;
Best Local Similarity 62.5%; Pred. No. 49;
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY      4 AVRLALSGSDGRVGS 19
Db      1761 AIQLSESGSDVDVGS 1776
      |:::| | | | | | |
RESULT 5
US-08-646-715-11
; Sequence 11, Application US/08646715
; Patent No. 5637686
; GENERAL INFORMATION:
; APPLICANT: Tjian, Robert
; APPLICANT: Comai, Lucio
; APPLICANT: Dynlacht, Brian D.
; APPLICANT: Hoey, Timothy
; APPLICANT: Ruppert, Siegfried
; APPLICANT: Tanese, Naoko
; APPLICANT: Wang, Edith
; APPLICANT: Weinzierl, Robert O.J.
; TITLE OF INVENTION: TATA-BINDING PROTEIN ASSOCIATED FACTORS,
; TITLE OF INVENTION: NUCLEIC ACIDS ENCODING TAFs AND METHODS OF USE
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOERBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/646,715
; FILING DATE: 09-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/188,582
; FILING DATE: 28-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-57650-2/AJT/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1893 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-646-715-11

Query Match          48.5%; Score 48; DB 1; Length 1893;
Best Local Similarity 62.5%; Pred. No. 49;
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY      4 AVRLALSGSDGRVGS 19

```

```

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31177
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31177

Query Match 45.5%; Score 45; DB 4; Length 432;
Best Local Similarity 62.5%; Pred. No. 28;
Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 LQRAVRLALSGDGRV 16
| : : : : : |
DB 199 LDRSARLSLGDPRV 214

RESULT 9
US-09-252-991A-32206
; Sequence 32206, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32206
; LENGTH: 660
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32206

Query Match 44.9%; Score 44.5; DB 4; Length 660;
Best Local Similarity 50.0%; Pred. No. 54;
Matches 11; Conservative 3; Mismatches 5; Indels 3; Gaps 1;

QY 2 QRAVRLALSGS---DGRVSGC 20
| : : : : : |
DB 326 QRSVRRARAGALQCGGRPHGC 347

RESULT 10
US-09-252-991A-25851
; Sequence 25851, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25851
; LENGTH: 546

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30646
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30646

Query Match 46.0%; Score 45.5; DB 4; Length 341;
Best Local Similarity 42.3%; Pred. No. 18;
Matches 11; Conservative 2; Mismatches 4; Indels 9; Gaps 1;

QY 4 AVRLALSGSDGR-----VGSGC 20
| : : : : : |
DB 1 AVRQAQGRDGRRRSCLWGQYLGNC 26

RESULT 7
US-09-252-991A-27887
; Sequence 27887, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27887
; LENGTH: 212
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27887

Query Match 45.5%; Score 45; DB 4; Length 212;
Best Local Similarity 72.7%; Pred. No. 13;
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 10 SGSDGRVGSGC 20
| : : : : : |
DB 133 SGSAARIGSGC 143

RESULT 8
US-09-252-991A-31177
; Sequence 31177, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
```

```
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25851

Query Match      44.4%; Score 44; DB 4; Length 546;
Best Local Similarity 58.8%; Pred. No. 53;
Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 3 RAVRLALSGDGRVSG 19
Db 195 RAHRLVLPAGGRRGG 211

RESULT 11
US-09-252-991A-18901
; Sequence 18901, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18901
; LENGTH: 410
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18901

Query Match      43.9%; Score 43.5; DB 4; Length 410;
Best Local Similarity 63.2%; Pred. No. 46;
Matches 12; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

QY 1 LQRAVRLALSGDGRVSG 19
Db 53 LQRVV-LALHGDDHVAVG 70

RESULT 12
US-09-252-991A-25050
; Sequence 25050, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25050
; LENGTH: 802
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25050

Query Match      42.9%; Score 42.5; DB 4; Length 802;
Best Local Similarity 46.2%; Pred. No. 1.4e+02;
Matches 12; Conservative 3; Mismatches 4; Indels 7; Gaps 1;

QY 1 LQRAVR-----LALSGDGRVSG 19
Db 663 LVRVARRTLPGAALTGATGRCGAG 688

; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23103
; Sequence 23103, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23103
; LENGTH: 134
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23103

Query Match      42.4%; Score 42; DB 4; Length 134;
Best Local Similarity 52.9%; Pred. No. 23;
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 4 AVRLALSGSDGRVSGC 20
Db 38 ARAAGGGRGARHSGC 54

RESULT 14
US-09-673-395A-235
; Sequence 235, Application US/09673395A
; Patent No. 6620923
; GENERAL INFORMATION:
; APPLICANT: SPECHT, THOMAS
; APPLICANT: HINZMANN, BERND
; APPLICANT: SCHMITT, ARMIN
; APPLICANT: PILARSKY, CHRISTIAN
; APPLICANT: DAHL, EDGAR
; APPLICANT: ROSENTHAL, ANDRE
; TITLE OF INVENTION: HUMAN NUCLEIC ACID SEQUENCES FROM UTERUS TUMOR TISSUE
; FILE REFERENCE: ALBRE-12
; CURRENT APPLICATION NUMBER: US/09/673,395A
; CURRENT FILING DATE: 2000-10-17
; NUMBER OF SEQ ID NOS: 637
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 235
; LENGTH: 187
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-673-395A-235

Query Match      42.4%; Score 42; DB 4; Length 187;
Best Local Similarity 45.5%; Pred. No. 33;
Matches 10; Conservative 3; Mismatches 5; Indels 4; Gaps 1;

QY 2 QRARLALSGSD-----GRVSG 19
Db 46 EKAIPALLEGKDLARARTGSG 67

RESULT 15
US-09-673-395A-571
; Sequence 571, Application US/09673395A
; Patent No. 6620923
; GENERAL INFORMATION:
; APPLICANT: SPECHT, THOMAS
; APPLICANT: HINZMANN, BERND
; APPLICANT: SCHMITT, ARMIN
```

; APPLICANT: PILARSKY, CHRISTIAN
; APPLICANT: DAHL, EDGAR
; APPLICANT: ROSENTHAL, ANDRE
; TITLE OF INVENTION: HUMAN NUCLEIC ACID SEQUENCES FROM UTERUS TUMOR TISSUE
; FILE REFERENCE: ALBRE-12
; CURRENT APPLICATION NUMBER: US/09/673,395A
; CURRENT FILING DATE: 2000-10-17
; NUMBER OF SEQ ID NOS: 637
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 571
; LENGTH: 195
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-673-395A-571

Query Match 42.4%; Score 42; DB 4; Length 195;
Best Local Similarity 45.5%; Pred. No. 35;
Matches 10; Conservative 3; Mismatches 5; Indels 4; Gaps 1;

QY 2 QRAVRLALSGSD----GRVGS 19
DB 46 EKAIPALLEGKOLLARARTGSG 67

RESULT 16
US-09-673-395A-324
; Sequence 324, Application US/09673395A
; Patent No. 6620923
; GENERAL INFORMATION:
; APPLICANT: SPECHT, THOMAS
; APPLICANT: HINZMANN, BERND
; APPLICANT: SCHMITT, ARMIN
; APPLICANT: PILARSKY, CHRISTIAN
; APPLICANT: DAHL, EDGAR
; APPLICANT: ROSENTHAL, ANDRE
; TITLE OF INVENTION: HUMAN NUCLEIC ACID SEQUENCES FROM UTERUS TUMOR TISSUE
; FILE REFERENCE: ALBRE-12
; CURRENT APPLICATION NUMBER: US/09/673,395A
; CURRENT FILING DATE: 2000-10-17
; NUMBER OF SEQ ID NOS: 637
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 324
; LENGTH: 224
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-673-395A-324

Query Match 42.4%; Score 42; DB 4; Length 224;
Best Local Similarity 45.5%; Pred. No. 41;
Matches 10; Conservative 3; Mismatches 5; Indels 4; Gaps 1;

QY 2 QRAVRLALSGSD----GRVGS 19
DB 46 EKAIPALLEGKOLLARARTGSG 67

RESULT 17
US-09-252-991A-19401
; Sequence 19401, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 19401

; LENGTH: 474
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19401

Query Match 42.4%; Score 42; DB 4; Length 474;
Best Local Similarity 52.6%; Pred. No. 94;
Matches 10; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 LQRAVRLALSGSDGRVGS 19
DB 58 LGRALRLAVSRQVGOTGAG 76

RESULT 18
US-09-176-657-1
; Sequence 1, Application US/09176657
; Patent No. 6020164
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Tang, Y. Tom
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Lu, Aina
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: HUMAN RNA BINDING PROTEINS
; FILE REFERENCE: PF-0611 US
; CURRENT APPLICATION NUMBER: US/09/176,657
; CURRENT FILING DATE: 1998-10-21
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PERL Program
; SEQ ID NO 1
; LENGTH: 547
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: 1229372
US-09-176-657-1

Query Match 42.4%; Score 42; DB 3; Length 547;
Best Local Similarity 45.5%; Pred. No. 1.1e+02;
Matches 10; Conservative 3; Mismatches 5; Indels 4; Gaps 1;

QY 2 QRAVRLALSGSD----GRVGS 19
DB 35 EKAIPALLEGKOLLARARTGSG 56

RESULT 19
US-09-421-299-1
; Sequence 1, Application US/09421299
; Patent No. 6524579
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Tang, Y. Tom
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Lu, Aina
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: HUMAN RNA BINDING PROTEINS
; FILE REFERENCE: PF-0611 US
; CURRENT APPLICATION NUMBER: US/09/421,299
; CURRENT FILING DATE: 1999-10-20
; EARLIER APPLICATION NUMBER: 09/176,657
; EARLIER FILING DATE: 1998-10-21
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PERL Program
; SEQ ID NO 1
; LENGTH: 547
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: 1229372

US-09-421-299-1

Query Match 42.4%; Score 42; DB 4; Length 547;
Best Local Similarity 45.5%; Pred. No. 1.1e+02;
Matches 10; Conservative 3; Mismatches 5; Indels 4; Gaps 1;

QY 2 QRAVRLALSGSD----GRVSG 19
Db 35 EKAIPALEGKOLLARARTSG 56

RESULT 20

US-09-205-258-264
; Sequence 264, Application US/09205258
; Patent No. 6525174
; GENERAL INFORMATION:
; APPLICANT: Young et al.
; TITLE OF INVENTION: 207 Human Secreted Proteins
; FILE REFERENCE: P2007P1
; CURRENT APPLICATION NUMBER: US/09/205,258
; EARLIER FILING DATE: 1998-12-04
; EARLIER APPLICATION NUMBER: PCT/US98/11422
; EARLIER FILING DATE: 1998-06-04
; EARLIER APPLICATION NUMBER: 60/048,895
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,375
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,881
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,890
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,896
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,020
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,876
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,895
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,884
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,894
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,971
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,964
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,882
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,899
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,893
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,900
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,901
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,892
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,915
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,019
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,970
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,972
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,916
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/049,373
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,875
; EARLIER FILING DATE: 1997-06-06

; EARLIER APPLICATION NUMBER: 60/049,374
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,917
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,949
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,883
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,897
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,898
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,962
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,963
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,877
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/048,878
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/070,923
; EARLIER FILING DATE: 1997-12-18
; EARLIER APPLICATION NUMBER: 60/092,921
; EARLIER FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: 60/094,657
; EARLIER FILING DATE: 1998-07-30
; NUMBER OF SEQ ID NOS: 1227
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 264
; LENGTH: 548
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (548)
; OTHER INFORMATION: Xaa equals stop translation
US-09-205-258-264

Query Match 42.4%; Score 42; DB 4; Length 548;
Best Local Similarity 45.5%; Pred. No. 1.1e+02;
Matches 10; Conservative 3; Mismatches 5; Indels 4; Gaps 1;

QY 2 QRAVRLALSGSD----GRVSG 19
Db 35 EKAIPALEGKOLLARARTSG 56

RESULT 21

US-08-492-459-10
; Sequence 10, Application US/08492459
; Patent No. 6015689
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: REGULATION OF AUROBASIDIN SENSITIVITY IN FUNGUS
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.4 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/492,459
; FILING DATE: June 20, 1995
; CLASSIFICATION: 435

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/243,403
;; FILING DATE: May 16, 1994
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Warren M. Cheek, Jr.
;; REGISTRATION NUMBER: 33,367
;; REFERENCE/DOCKET NUMBER:
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 202-371-8850
;; TELEFAX:
;; TELEX:

;; INFORMATION FOR SEQ ID NO: 10:

;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1477
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
US-08-492-459-10

Query Match 42.4%; Score 42; DB 3; Length 1477;
Best Local Similarity 42.1%; Pred. No. 3.4e+02;
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 1 LQRAVRLALSGSDGRVGS 19
Db 772 IERASRVVLGTGQVDIG 790

RESULT 22

US-08-423-752-10
;; Sequence 10, Application US/08423752
;; Patent No. 6022949

;; GENERAL INFORMATION:

;; APPLICANT: Takashi OKADO et al.
;; TITLE OF INVENTION: A GENE CODING FOR A PROTEIN REGULATING
;; TITLE OF INVENTION: AUROBASIDIN SENSITIVITY
;; NUMBER OF SEQUENCES: 22

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: Wenderoth, Lind & Ponack
;; STREET: 805 Fifteenth Street, N.W., #700
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: U.S.A.
;; ZIP: 20005

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: MS-DOS
;; SOFTWARE: Wordperfect 5.1

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/423,752
;; FILING DATE: April 18, 1995
;; CLASSIFICATION: 435

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: 08/243,403
;; FILING DATE: May 16, 1994

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Warren M. Cheek, Jr.
;; REGISTRATION NUMBER: 33,367
;; REFERENCE/DOCKET NUMBER:
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 202-371-8850
;; TELEFAX:

;; TELEX:

;; INFORMATION FOR SEQ ID NO: 10:

;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1477
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
US-08-423-752-10

Query Match 42.4%; Score 42; DB 3; Length 1477;
Best Local Similarity 42.1%; Pred. No. 3.4e+02;
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 1 LQRAVRLALSGSDGRVGS 19
Db 772 IERASRVVLGTGQVDIG 790

RESULT 23

US-08-945-994-3

;; Sequence 3, Application US/08945994
;; Patent No. 6043051

;; GENERAL INFORMATION:

;; APPLICANT: Takashi OKADO et al.
;; TITLE OF INVENTION: PROMOTER
;; NUMBER OF SEQUENCES: 8
;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
;; STREET: 2033 K Street, N.W., #800
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: U.S.A.
;; ZIP: 20006

;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: MS-DOS
;; SOFTWARE: Wordperfect 5.1

;; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/945,994
;; FILING DATE: No. 6043051ember 6, 1997
;; CLASSIFICATION: 435

;; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER:

;; FILING DATE:

;; ATTORNEY/AGENT INFORMATION:

;; NAME: Warren M. Cheek, Jr.

;; REGISTRATION NUMBER: 33,367

;; REFERENCE/DOCKET NUMBER:

;; TELECOMMUNICATION INFORMATION:

;; TELEPHONE: 202-721-8200

;; TELEFAX:

;; TELEX:

;; INFORMATION FOR SEQ ID NO: 3:

;; SEQUENCE CHARACTERISTICS:

;; LENGTH: 1477

;; TYPE: amino acid

;; STRANDEDNESS: single

;; TOPOLOGY: linear

;; MOLECULE TYPE: peptide

US-08-945-994-3

Query Match 42.4%; Score 42; DB 3; Length 1477;
Best Local Similarity 42.1%; Pred. No. 3.4e+02;
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 1 LQRAVRLALSGSDGRVGS 19
Db 772 IERASRVVLGTGQVDIG 790

RESULT 24

US-08-716-873-24

;; Sequence 24, Application US/08716873

;; Patent No. 6194166

;; GENERAL INFORMATION:

;; APPLICANT: Takashi OKADO et al.

;; TITLE OF INVENTION: GENE REGULATING AUROBASIDIN SENSITIVITY

;; NUMBER OF SEQUENCES: 50

;; CORRESPONDENCE ADDRESS:

;; ADDRESSEE: Wenderoth, Lind & Ponack

STREET: 805 Fifteenth Street, N.W., #700
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/716,873
FILING DATE: September 20, 1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8850
TELEFAX:
TELEX:
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 1477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-716-873-24

Query Match 42.4%; Score 42; DB 3; Length 1477;
Best Local Similarity 42.1%; Pred. No. 3.4e+02;
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 1 LQRAVRLALSGSDGRVGS 19
Db 772 IERASRVIVLTGQGVIG 790

RESULT 25
US-09-368-431-24
Sequence 24, Application US/09368431
Patent No. 6294651
GENERAL INFORMATION:
APPLICANT: Takashi OKADO et al.
TITLE OF INVENTION: GENE REGULATING AUREOBASIDIN SENSITIVITY
TITLE OF INVENTION: (AS AMENDED)
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: WENDEROTH, LIND & PONACK, L.L.P.
STREET: 2033 K Street, N.W., #800
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/368,431
FILING DATE: August 5, 1999
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/716,873
FILING DATE: September 20, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.

REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX:
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 1477 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-368-431-24

Query Match 42.4%; Score 42; DB 3; Length 1477;
Best Local Similarity 42.1%; Pred. No. 3.4e+02;
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 1 LQRAVRLALSGSDGRVGS 19
Db 772 IERASRVIVLTGQGVIG 790

RESULT 26
US-09-414-006-10
Sequence 10, Application US/09414006
Patent No. 6348577
GENERAL INFORMATION:
APPLICANT: Takashi OKADO et al.
TITLE OF INVENTION: REGULATION OF AUREOBASIDIN SENSITIVITY (AS AMENDED)
NUMBER OF SEQUENCES: 36
CORRESPONDENCE ADDRESS:
ADDRESSEE: WENDEROTH, LIND & PONACK, L.L.P.
STREET: 2033 K Street, N.W., #800
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.4 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/414,006
FILING DATE: October 7, 1999
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/492,459
FILING DATE: June 20, 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/243,403
FILING DATE: May 16, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX:
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1477
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-414-006-10

Query Match 42.4%; Score 42; DB 4; Length 1477;

Best Local Similarity 42.1%; Pred. No. 3.4e+02; Indels 0; Gaps 0;
Matches 8; Conservative 6; Mismatches 5;

QY 1 LORAVRLALSGSDGRVSG 19
Db 772 IERASRVVLGTDGQVDIG 790

RESULT 27

US-09-447-223-10
; Sequence 10, Application US/09447223
; Patent No. 6432664
; GENERAL INFORMATION:
; APPLICANT: Takashi OKADO et al.
; TITLE OF INVENTION: A GENE CODING FOR A PROTEIN REGULATING
; ;
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; FILING DATE: 23-No. 6432664-1999
; CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/243,403
; FILING DATE: May 16, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER: <Unknown>
; TELEPHONE: 202-371-8850
; TELEFAX: <Unknown>
; TELEX: <Unknown>

INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1477
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-447-223-10

Query Match 42.4%; Score 42; DB 4; Length 1477;
Best Local Similarity 42.1%; Pred. No. 3.4e+02; Indels 0; Gaps 0;
Matches 8; Conservative 6; Mismatches 5;

QY 1 LORAVRLALSGSDGRVSG 19
Db 772 IERASRVVLGTDGQVDIG 790

RESULT 28

US-09-543-681A-5986
; Sequence 5986, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A

; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 5986
; LENGTH: 1495
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-5986

Query Match 41.9%; Score 41.5; DB 4; Length 1495;
Best Local Similarity 34.5%; Pred. No. 4.1e+02;
Matches 10; Conservative 3; Mismatches 5; Indels 11; Gaps 1;

QY 3 RAVRLALSG-----SDGRVSGC 20
Db 37 KIVRNAIHLARMQHRGAILSDGKTGDGC 65

RESULT 29

US-08-362-670B-30
; Sequence 30, Application US/08362670B
; Patent No. 5658882
; GENERAL INFORMATION:
; APPLICANT: Celeste, Anthony J.
; APPLICANT: Wozney, John
; APPLICANT: Rosen, Vicki A.
; APPLICANT: Wolfman, Neil
; APPLICANT: Thomsen, Gerald H.
; APPLICANT: Melton, Douglas A.
; TITLE OF INVENTION: TENDON-INDUCING COMPOSITIONS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENETICS INSTITUTE, INC.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362,670B
; FILING DATE: December 22, 1994
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Lazar, Steven R.
; REGISTRATION NUMBER: 32,618
; REFERENCE/DOCKET NUMBER: 5202-D
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617 498-8260
; TELEFAX: 617 876-5851
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 240 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-362-670B-30

Query Match 41.4%; Score 41; DB 1; Length 240;
Best Local Similarity 50.0%; Pred. No. 64;
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 2 QRAVRLALSGSDGRVSG 19
Db 90 RRRQRTALAGTGGXXGSG 107

RESULT 30

US-08-333-576C-30
 ; Sequence 30, Application US/08333576C
 ; Patent No. 6027919
 ; GENERAL INFORMATION:
 ; APPLICANT: Celeste, Anthony J.
 ; APPLICANT: Wozney, John
 ; APPLICANT: Rosen, Vicki A.
 ; APPLICANT: Wolfman, Neil
 ; APPLICANT: Thomsen, Gerald H.
 ; APPLICANT: Melton, Douglas A.
 ; TITLE OF INVENTION: TENDON-INDUCING COMPOSITIONS
 ; NUMBER OF SEQUENCES: 37
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: GENETICS INSTITUTE, INC.
 ; STREET: 87 CambridgePark Drive
 ; CITY: Cambridge
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02140
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/333,576C
 ; FILING DATE: No. 6027919ember 2, 1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Lazar, Steven R.
 ; REGISTRATION NUMBER: 32,618
 ; REFERENCE/DOCKET NUMBER: 5202-B
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617 498-8260
 ; TELEFAX: 617 876-5851
 ; INFORMATION FOR SEQ ID NO: 30:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 240 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-333-576C-30

Query Match 41.4%; Score 41; DB 3; Length 240;
 Best Local Similarity 50.0%; Pred. NO. 64;
 Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
 Qy 2 QRAVRLALSGSDGRVCSG 19
 Db 90 RRRQRTALAGTRGXGSG 107

Search completed: May 13, 2004, 06:56:40
 Job time : 15 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2004, 06:55:22 ; Search time 214 Seconds
(without alignments)
26.006 Million cell updates/sec

Title: 09549186-9

Perfect score: 99

Sequence: 1 LQRAVRLALSGSDGRVSGSC 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1145568 seqs, 278261457 residues

Total number of hits satisfying chosen parameters: 1145568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Published Applications_AA.*

1: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/prodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/prodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/prodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/prodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/prodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/prodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/prodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	52	52.5	162	12	US-10-425-114-50038, A
2	52	52.5	233	12	US-10-425-114-72819, A
3	52	52.5	247	12	US-10-425-114-38641, A
4	48	48.5	898	14	US-10-043-487-277
5	47	47.5	273	15	US-10-369-493-1170, App
6	46.5	47.0	300	14	US-10-306-762-80
7	45	45.5	359	9	US-09-815-242-5116
8	45	45.5	359	12	US-10-282-122A-43489
9	45	45.5	592	16	US-10-389-566-1504
10	45	45.5	892	16	US-10-293-965-9
11	45	45.5	1399	14	US-10-156-761-14753
12	44	44.4	49	12	US-10-424-599-268979
13	44	44.4	243	12	US-10-412-699B-1447
14	44	44.4	243	15	US-10-310-154-642
15	44	44.4	243	15	US-10-374-780A-1302

44	44.4	296	12	US-10-425-114-42184	Sequence 42184, A
44	44.4	478	15	US-10-369-493-16830	Sequence 16830, A
43	43.4	215	14	US-10-156-761-12749	Sequence 12749, A
19	43	257	15	US-10-369-493-17070	Sequence 17070, A
20	43.4	318	15	US-10-369-493-4692	Sequence 4692, Ap
21	43.4	585	16	US-10-389-566-1035	Sequence 1035, Ap
22	43	589	16	US-10-389-566-1753	Sequence 1753, Ap
23	43.4	760	15	US-10-369-493-9961	Sequence 9961, Ap
24	43.4	943	15	US-10-369-493-20410	Sequence 20410, A
25	43.4	964	15	US-10-369-493-1347	Sequence 1347, Ap
26	43.4	964	16	US-10-389-566-1092	Sequence 1092, Ap
27	43.4	1013	16	US-10-389-566-1473	Sequence 1473, Ap
28	43.4	1017	15	US-10-369-493-21692	Sequence 21692, A
29	43.4	1017	16	US-10-389-566-2218	Sequence 2218, Ap
30	42.5	229	9	US-09-738-626-6612	Sequence 6612, Ap
31	42.5	986	12	US-10-282-122A-42808	Sequence 42808, A
32	42.4	85	9	US-09-864-761-42554	Sequence 42554, A
33	42.4	135	9	US-09-925-302-492	Sequence 492, App
34	42.4	135	12	US-09-925-302-492	Sequence 492, App
35	42.4	157	12	US-10-425-114-53027	Sequence 53027, A
36	42.4	310	15	US-10-240-145-164	Sequence 164, App
37	42.4	383	14	US-10-156-761-12847	Sequence 12847, A
38	42.4	409	12	US-10-282-122A-72287	Sequence 72287, A
39	42.4	423	12	US-10-424-589-144979	Sequence 144979, A
40	42.4	547	14	US-10-359-385-1	Sequence 1, Appli
41	42.4	548	10	US-09-933-767-264	Sequence 264, App
42	42.4	548	12	US-10-004-860-264	Sequence 264, App
43	42.4	548	14	US-10-023-282-264	Sequence 264, App
44	42.4	577	12	US-10-072-012-565	Sequence 565, App
45	42.4	600	12	US-10-092-900A-216	Sequence 216, App
46	42.4	922	15	US-10-369-493-5105	Sequence 5105, Ap
47	42.4	1367	15	US-10-369-493-5508	Sequence 5508, Ap
48	42.4	1367	15	US-10-369-493-5509	Sequence 5509, Ap
49	42.4	1477	10	US-09-951-217-24	Sequence 24, Appl
50	42.4	1477	12	US-10-388-215-54	Sequence 54, Appl
51	42.4	1477	15	US-10-369-493-22009	Sequence 22009, A
52	42.4	1973	15	US-10-369-493-2380	Sequence 2380, Ap
53	42.4	19695	15	US-10-084-846A-3	Sequence 3, Appli
54	41.5	78	10	US-08-764-872-305	Sequence 305, App
55	41.5	787	12	US-10-282-122A-45297	Sequence 45297, A
56	41.5	1485	12	US-10-282-122A-69097	Sequence 69097, A
57	41.4	43	10	US-08-899-495-37	Sequence 97, Appl
58	41.4	79	12	US-10-424-599-206798	Sequence 206798, A
59	41.4	152	12	US-10-425-114-48751	Sequence 48751, A
60	41.4	162	15	US-10-364-743-68	Sequence 68, Appl
61	41.4	166	15	US-10-364-743-32	Sequence 32, Appl
62	41.4	240	9	US-09-945-182-30	Sequence 30, Appl
63	41.4	261	15	US-10-369-493-8867	Sequence 8867, Ap
64	41.4	269	12	US-10-282-122A-77449	Sequence 77449, A
65	41.4	417	12	US-10-425-114-56059	Sequence 56059, A
66	41.4	531	12	US-10-282-122A-66939	Sequence 66939, A
67	41.4	2644	14	US-10-300-453A-39	Sequence 39, Appl
68	41.4	2654	14	US-10-227-610-2	Sequence 2, Appli
69	41.4	3564	14	US-10-156-761-7964	Sequence 7964, Ap
70	41.4	3619	12	US-10-282-122A-67768	Sequence 67768, A
71	40.4	31	9	US-09-864-761-42530	Sequence 42530, A
72	40.4	68	12	US-10-424-599-227312	Sequence 227312, A
73	40.4	77	14	US-10-106-698-6588	Sequence 6588, Ap
74	40.4	165	14	US-10-040-895-16	Sequence 16, Appl
75	40.4	247	9	US-09-738-626-5108	Sequence 5108, Ap
76	40.4	264	15	US-10-369-493-12326	Sequence 12326, A
77	40.4	273	9	US-09-815-242-10017	Sequence 10017, A
78	40.4	273	9	US-09-815-242-14012	Sequence 14012, A
79	40.4	273	12	US-10-282-122A-55751	Sequence 55751, A
80	40.4	273	12	US-10-282-122A-56401	Sequence 56401, A
81	40.4	273	12	US-10-282-122A-60211	Sequence 60211, A
82	40.4	273	12	US-10-282-122A-74943	Sequence 74943, A
83	40.4	273	12	US-10-282-122A-75576	Sequence 75576, A
84	40.4	273	15	US-10-369-493-690	Sequence 690, App
85	40.4	279	12	US-10-282-122A-60685	Sequence 60685, A
86	40.4	289	12	US-10-282-122A-59106	Sequence 59106, A
87	40.4	299	15	US-10-369-493-11875	Sequence 11875, A
88	40.4	330	15	US-10-369-493-23413	Sequence 23413, A

Sequence 1315, Ap
Sequence 20380, A
Sequence 12263, A
Sequence 224926, A
Sequence 62711, A
Sequence 720, App
Sequence 762, App
Sequence 2526, App
Sequence 45413, A
Sequence 78515, A
Sequence 56580, A
Sequence 69032, A

ALIGNMENTS

RESULT 1

US-10-425-114-50038
; Sequence 50038, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 50038
; LENGTH: 162
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3592-085-D2_FLI.pep
US-10-425-114-50038

Query Match 52.5%; Score 52; DB 12; Length 162;
Best Local Similarity 64.3%; Pred. No. 4.6;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 7 LALSGSDGRVSGC 20

Db 39 VALAGSDGRVSGC 52

RESULT 2

US-10-425-114-72819
; Sequence 72819, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 72819
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3592-085-D2_FLI.pep
US-10-425-114-72819

Query Match 52.5%; Score 52; DB 12; Length 233;
Best Local Similarity 64.3%; Pred. No. 4.6;
Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OTHER INFORMATION: Clone ID: LIB4721-024-H12_FLI.pep
US-10-425-114-72819

Query Match 52.5%; Score 52; DB 12; Length 233;
Best Local Similarity 71.4%; Pred. No. 6.8;
Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 7 LALSGSDGRVSGC 20

Db 96 VALAGSDGRVSGC 109

RESULT 3

US-10-425-114-38641
; Sequence 38641, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 38641
; LENGTH: 247
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700350355_FLI.pep
US-10-425-114-38641

Query Match 52.5%; Score 52; DB 12; Length 247;
Best Local Similarity 71.4%; Pred. No. 7.2;
Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 7 LALSGSDGRVSGC 20

Db 108 VALAGSDGRVSGC 121

RESULT 4

US-10-043-487-277
; Sequence 277, Application US/10043487
; Publication No. US20030055220A1
; GENERAL INFORMATION:
; APPLICANT: HYBRIGENICS
; APPLICANT: Pierre, LEBRAIN
; TITLE OF INVENTION: Protein-protein interactions between Shigella Flexneri polypeptides
; TITLE OF INVENTION: mammalian polypeptides
; FILE REFERENCE: B4778A
; CURRENT APPLICATION NUMBER: US/10/043,487
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/261,130
; PRIOR FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 561
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 277
; LENGTH: 898
; TYPE: PRT
; ORGANISM: Shigella Flexneri
; OTHER INFORMATION: Clone ID: 700350355_FLI.pep
US-10-043-487-277

Query Match 48.5%; Score 48; DB 14; Length 898;
Best Local Similarity 62.5%; Pred. No. 1.1e+02;
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 AVRLALSGSDGRVSG 19

```

Db      766 AIOLESSESDSDVGS 781
      |:::|:|||||:||||
RESULT 5
US-10-369-493-1170
; Sequence 1170, Application US/10369493
; Publication No. US2003033675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1170
; LENGTH: 273
; TYPE: PRT
; ORGANISM: Methanobacterium thermoautotrophicum
US-10-369-493-1170
Query Match      47.5%; Score 47; DB 15; Length 273;
Best Local Similarity 53.3%; Pred. No. 46;
Matches 8; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY      5 VRLALSGSDGRVGS 19
      |:::|:|||||:||||
Db      2 IRVATGACRMVGS 16
      |:::|:|||||:||||
RESULT 6
US-10-306-762-80
; Sequence 80, Application US/10306762
; Publication No. US2003018720A1
; GENERAL INFORMATION:
; APPLICANT: Park, Frances
; APPLICANT: Gajiwala, Ketan S.
; APPLICANT: Buchanan, Sean Grant
; APPLICANT: Sauder, J. Michael
; TITLE OF INVENTION: CRYSTALS AND STRUCTURES OF A FLAVIN
; FILE REFERENCE: MONONUCLEOTIDE BINDING PROTEIN (FMNBP)
; CURRENT APPLICATION NUMBER: US/10/306,762
; CURRENT FILING DATE: 2003-04-16
; PRIOR APPLICATION NUMBER: US 60/334,132
; PRIOR FILING DATE: 2001-11-28
; NUMBER OF SEQ ID NOS: 242
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 80
; LENGTH: 300
; TYPE: PRT
; ORGANISM: G. diazotrophicus (20257536)
US-10-306-762-80
Query Match      47.0%; Score 46.5; DB 14; Length 300;
Best Local Similarity 66.7%; Pred. No. 60;
Matches 10; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

QY      7 LALSGSDG-RVGS 20
      |:::|:|||||:||||
Db      194 LALSGADGVMIGRC 208
      |:::|:|||||:||||
RESULT 7
US-09-815-242-5116
; Sequence 5116, Application US/09815242
```

```

; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-03-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5116
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-5116
Query Match      45.5%; Score 45; DB 9; Length 359;
Best Local Similarity 62.5%; Pred. No. 1.2e+02;
Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY      1 LORAVRLALSGSDGRV 16
      |:::|:|||||:||||
Db      126 LDRSARLSLSDPRV 141
      |:::|:|||||:||||
RESULT 8
US-10-282-122A-43489
; Sequence 43489, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
```

```
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 43489
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-43489

Query Match      45.5%; Score 45; DB 12; Length 359;
Best Local Similarity 62.5%; Pred. No. 1.2e+02;
Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 LQRAVRLALSGSDGRV 16
   |||:|||||
Db 126 LDRSARLSLSDPRV 141

RESULT 9
US-10-389-566-1504
; Sequence 1504, Application US/10389566
; Publication No. US20040025202A1
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: US/10/389,566
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1504
; LENGTH: 592
; TYPE: PRT
; ORGANISM: Methanopyrus kandleri
US-10-389-566-1504

Query Match      45.5%; Score 45; DB 16; Length 592;
Best Local Similarity 58.8%; Pred. No. 2.1e+02;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 QRAVRLALSGSDGRV 18
   :|||:|||||
Db 371 ERAGRVVCLGSDDRV 387

RESULT 10
US-10-293-965-9
; Sequence 9, Application US/10293965
; Publication No. US20040025196A1
; GENERAL INFORMATION:
; APPLICANT: ALROY et al
; TITLE OF INVENTION: POSH NUCLEIC ACIDS, POLYPEPTIDES AND RELATED METHODS
```

```
; FILE REFERENCE: PROL-P01-010
; CURRENT APPLICATION NUMBER: US/10/293,965
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: 60/345846
; PRIOR FILING DATE: 2001-11-09
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 892
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-293-965-9

Query Match      45.5%; Score 45; DB 16; Length 892;
Best Local Similarity 65.0%; Pred. No. 3.2e+02;
Matches 13; Conservative 0; Mismatches 5; Indels 2; Gaps 1;

QY 1 LQRAV--RLALSGSDGRV 18
   |||:|||||
Db 759 LQGAVGPELPLGGSHGRV 778

RESULT 11
US-10-156-761-14753
; Sequence 14753, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 14753
; LENGTH: 1399
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-14753

Query Match      45.5%; Score 45; DB 14; Length 1399;
Best Local Similarity 55.6%; Pred. No. 5.1e+02;
Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 LQRAVRLALSGSDGRV 18
   |||:|||||
Db 901 LRRRVRAVLSAADGRV 918

RESULT 12
US-10-424-599-268979
; Sequence 268979, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
```

; SEQ ID NO 268979

; LENGTH: 49

; TYPE: PRT

; ORGANISM: Glycine max

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT_MRT3847_8490C.1.pep

US-10-424-599-268979

Query Match 44.4%; Score 44; DB 12; Length 49;

Best Local Similarity 56.2%; Pred. No. 22;

Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 QRAVLALSGSGRVG 17

DB 34 QPAITVLSHADGRVG 49

RESULT 13

US-10-412-699B-1447

; Sequence 1447, Application US/10412699B

; Publication No. US20040045049A1

; GENERAL INFORMATION:

; APPLICANT: Mendel Biotechnology, Inc.

; APPLICANT: Zhang, James

; APPLICANT: Fromm, Michael E.

; APPLICANT: Heard, Jacqueline E.

; APPLICANT: Riechmann, Jose Luis

; APPLICANT: Adam, Luc J.

; APPLICANT: Broun, Pierre B.

; APPLICANT: Pineda, Omaira

; APPLICANT: Reuber, T. Lynne

; APPLICANT: Keddle, James S.

; APPLICANT: Yu, Guo-Liang

; APPLICANT: Jiang, Cai-Zhong

; APPLICANT: Samaha, Raymond R.

; APPLICANT: Pilgrim, Marsha L.

; APPLICANT: Creelman, Robert A.

; APPLICANT: DuBell, Arnold N.

; APPLICANT: Ratcliffe, Oliver

; APPLICANT: Kumimoto, Roderick

; APPLICANT: Sherman, Bradley K.

; TITLE OF INVENTION: Polynucleotides and Polypeptides in Plants

; FILE REFERENCE: MBI-0048CIP

; CURRENT APPLICATION NUMBER: US/10/412,699B

; CURRENT FILING DATE: 2003-04-10

; PRIOR APPLICATION NUMBER: 09/394,519

; PRIOR FILING DATE: 1999-09-13

; PRIOR APPLICATION NUMBER: 09/489,376

; PRIOR FILING DATE: 2000-01-21

; PRIOR APPLICATION NUMBER: 09/506,720

; PRIOR FILING DATE: 2000-02-17

; PRIOR APPLICATION NUMBER: 09/533,030

; PRIOR FILING DATE: 2000-03-22

; PRIOR APPLICATION NUMBER: 09/533,392

; PRIOR FILING DATE: 2000-03-22

; PRIOR APPLICATION NUMBER: 09/533,029

; PRIOR FILING DATE: 2000-03-22

; PRIOR APPLICATION NUMBER: 09/532,591

; PRIOR FILING DATE: 2000-03-22

; PRIOR APPLICATION NUMBER: 09/533,648

; PRIOR FILING DATE: 2000-03-22

; PRIOR APPLICATION NUMBER: 09/713,994

; PRIOR FILING DATE: 2000-11-16

; PRIOR APPLICATION NUMBER: 09/819,142

; PRIOR FILING DATE: 2001-03-27

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 2011

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 1447

; LENGTH: 243

; TYPE: PRT

; ORGANISM: Oryza sativa

US-10-412-699B-1447

Query Match

Best Local Similarity 44.4%; Score 44; DB 12; Length 243;

Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 8 ALSGSDGRVGG 19

DB 78 AVAGRDGRVGG 89

RESULT 14

US-10-310-154-642

; Sequence 642, Application US/10310154

; Publication No. US20030233670A1

; GENERAL INFORMATION:

; APPLICANT: Edgerton, Michael D

; APPLICANT: Chomet, Paul S.

; APPLICANT: Adams, Thomas H

; APPLICANT: Ruff, Thomas G.

; APPLICANT: Agarwal, Ameeta K.

; APPLICANT: Ahrens, Jeffrey E.

; APPLICANT: Ball, James A.

; APPLICANT: Banu, G

; APPLICANT: Bell, Brin

; APPLICANT: Boddupalli, Raghava

; APPLICANT: Deikman, Jill

; APPLICANT: Deng, Molian

; APPLICANT: Dong, Jinzhuo

; APPLICANT: Duff, Stephen M.

; APPLICANT: Galligan, Meghan M.

; APPLICANT: Hinchey, Brenda S.

; APPLICANT: Huang, Shihshieh

; APPLICANT: Johnson, G. Richard

; APPLICANT: Jung, Vincent

; APPLICANT: Kretzmer, Keith A

; APPLICANT: Laccetti, Lucille B.

; APPLICANT: Lai, Chao-Qiang

; APPLICANT: Lee, Gary

; APPLICANT: Lin, Jie-Yi

; APPLICANT: Liu, Jingdong

; APPLICANT: Lu, Bin

; APPLICANT: Luethy, Michael M.

; APPLICANT: Lund, Adrian

; APPLICANT: Madson, Linda L.

; APPLICANT: Malloy, Kathleen A.

; APPLICANT: McKiel, Christine L.

; APPLICANT: Miller, Philip W.

; APPLICANT: Padmavathi, Manchikanti

; APPLICANT: Parnell, Laurence D.

; APPLICANT: Start, William G.

; APPLICANT: Tennesen, Dan

; APPLICANT: Vidya, K.R.

; APPLICANT: Wang, Haiyun

; APPLICANT: Xin, Zhanqun

; APPLICANT: Xu, Nanfei

; APPLICANT: Yang, Chunzhi

; APPLICANT: Zeng, Xiaoping

; APPLICANT: Zhang, Qiang

; APPLICANT: Zhao, Yajuan

; APPLICANT: Zhou, Li

; TITLE OF INVENTION: Gene Sequences and Uses Thereof in Plants

; FILE REFERENCE: 38-15(52796)B

; CURRENT APPLICATION NUMBER: US/10/310,154

; CURRENT FILING DATE: 2002-12-04

; PRIOR APPLICATION NUMBER: 60/337,358

; PRIOR FILING DATE: 2001-12-04

; NUMBER OF SEQ ID NOS: 736

; SEQ ID NO 642

; LENGTH: 243

; TYPE: PRT

; ORGANISM: Oryza sativa

US-10-310-154-642


```
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 12749
; LENGTH: 215
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-12749

Query Match      43.4%; Score 43; DB 14; Length 215;
Best Local Similarity 57.1%; Pred. No. 1.4e+02;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 5 VRLALSGSDGRVGS 18
Db 43 IRLARGCDRTGT 56

RESULT 19
US-10-369-493-17070
; Sequence 17070, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 17070
; LENGTH: 257
; TYPE: PRT
; ORGANISM: Caulobacter crescentus
US-10-369-493-17070

Query Match      43.4%; Score 43; DB 15; Length 257;
Best Local Similarity 35.3%; Pred. No. 1.7e+02;
Matches 6; Conservative 9; Mismatches 2; Indels 0; Gaps 0;

QY 1 LQRAVRLALSGSDGRVGS 17
Db 1 MSQPVKIIATAGNRWG 17

RESULT 20
US-10-369-493-4692
; Sequence 4692, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
```

```
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 4692
; LENGTH: 318
; TYPE: PRT
; ORGANISM: Burkholderia fungorum
US-10-369-493-4692

Query Match      43.4%; Score 43; DB 15; Length 318;
Best Local Similarity 46.7%; Pred. No. 2.2e+02;
Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 6 RLALSGSDGRVGS 20
Db 139 RVVILGASGGVGTGC 153

RESULT 21
US-10-389-566-1035
; Sequence 1035, Application US/10389566
; Publication No. US20040025202A1
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: US/10/389,566
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1035
; LENGTH: 585
; TYPE: PRT
; ORGANISM: Thermococcus sp.
US-10-389-566-1035

Query Match      43.4%; Score 43; DB 16; Length 585;
Best Local Similarity 58.8%; Pred. No. 4.1e+02;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 QRAVRLALSGSDGRVGS 18
Db 366 ERAGRVTILGSDERVGS 382

RESULT 22
US-10-389-566-1753
; Sequence 1753, Application US/10389566
; Publication No. US20040025202A1
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: US/10/389,566
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
```

```
; SEQ ID NO 1753
; LENGTH: 589
; TYPE: PRT
; ORGANISM: Clostridium thermocellum
US-10-389-566-1753

Query Match      43.4%; Score 43; DB 16; Length 589;
Best Local Similarity 52.8%; Pred. No. 4.1e+02;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 QRAVRLALSGSDGRVGS 18
   :|||:|||||:
Db 367 ERAGRVVCLGSDGREGA 383

RESULT 23
US-10-369-493-9961
; Sequence 9961, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 9961
; LENGTH: 760
; TYPE: PRT
; ORGANISM: magnetite-containing magnetic coccus
US-10-369-493-9961

Query Match      43.4%; Score 43; DB 15; Length 760;
Best Local Similarity 40.0%; Pred. No. 5.4e+02;
Matches 8; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 1 LQRAVRLALSGSDGRVGS 20
   |::|||::|:
Db 456 LDKNLRLSLFGQGDGAIKTC 475

RESULT 24
US-10-369-493-20410
; Sequence 20410, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20410
; LENGTH: 943
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(943)

; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-20410

Query Match      43.4%; Score 43; DB 15; Length 943;
Best Local Similarity 58.8%; Pred. No. 6.7e+02;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 QRAVRLALSGSDGRVGS 18
   :|||:|||||:
Db 751 ERAGRVVTLGSDYRVGS 767

RESULT 25
US-10-369-493-1347
; Sequence 1347, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 1347
; LENGTH: 964
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii
US-10-369-493-1347

Query Match      43.4%; Score 43; DB 15; Length 964;
Best Local Similarity 58.8%; Pred. No. 6.9e+02;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 QRAVRLALSGSDGRVGS 18
   :|||:|||||:
Db 745 ERAGRVVTLGSDYRVGS 761

RESULT 26
US-10-389-566-1092
; Sequence 1092, Application US/10389566
; Publication No. US20040025202A1
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: US/10/389,566
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1092
; LENGTH: 964
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii
US-10-389-566-1092

Query Match      43.4%; Score 43; DB 16; Length 964;
Best Local Similarity 58.8%; Pred. No. 6.9e+02;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
```

QY 2 QRAVRLALSGSDGRVGS 18
: ||| : ||| ||||
Db 745 ERAGRVVTILGSDYRVGS 761

RESULT 27

US-10-389-566-1473
; Sequence 1473, Application US/10389566
; Publication No. US20040025202A1
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: US/10/389,566
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1473
; LENGTH: 1013
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
US-10-389-566-1473

Query Match 43.4%; Score 43; DB 16; Length 1013;
Best Local Similarity 58.8%; Pred. No. 7.2e+02;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 QRAVRLALSGSDGRVGS 18
: ||| : ||| ||||
Db 794 ERAGRVVTILGSDYRVGS 810

RESULT 28

US-10-369-493-21692
; Sequence 21692, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 21692
; LENGTH: 1017
; TYPE: PRT
; ORGANISM: Pyrococcus abyssi
US-10-369-493-21692

Query Match 43.4%; Score 43; DB 15; Length 1017;
Best Local Similarity 58.8%; Pred. No. 7.3e+02;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 QRAVRLALSGSDGRVGS 18
: ||| : ||| ||||
Db 798 ERAGRVVTILGSDYRVGS 814

RESULT 29

US-10-389-566-2218
; Sequence 2218, Application US/10389566
; Publication No. US20040025202A1
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: US/10/389,566
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2218
; LENGTH: 1017
; TYPE: PRT
; ORGANISM: Pyrococcus abyssi
US-10-389-566-2218

Query Match 43.4%; Score 43; DB 16; Length 1017;
Best Local Similarity 58.8%; Pred. No. 7.3e+02;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 QRAVRLALSGSDGRVGS 18
: ||| : ||| ||||
Db 798 ERAGRVVTILGSDYRVGS 814

RESULT 30

US-09-738-626-6612
; Sequence 6612, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6612
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6612

Query Match 42.9%; Score 42.5; DB 9; Length 229;
Best Local Similarity 47.6%; Pred. No. 1.8e+02;
Matches 10; Conservative 3; Mismatches 5; Indels 3; Gaps 1;

QY 2 QRAVRLALSG---SDGRVGS 19
: ||| : ||| : ||| : |||

Db 168 QRATQAAFTGYLTGDGRLGCG 188

Search completed: May 13, 2004, 07:17:03
Job time : 214 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2004, 06:49:47 ; Search time 12 Seconds
(without alignments)
160.319 Million cell updates/sec

Title: 09549186-9
Perfect score: 99
Sequence: 1 LQRAVRLALSGDGRVSGC 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : PIR 78:.*
1: piri:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49	49.5	575	D69512	acetolactate synth
2	48	48.5	1865	1 I48155	transcription init
3	48	48.5	1893	1 A40262	transcription init
4	47	47.5	273	H69206	dihydrodipicolinat
5	46	46.5	352	2 A81682	chitinases homolog
6	46	46.5	352	2 AC1310	chitinases homolog
7	45	45.5	359	2 E83312	hypothetical prote
8	45	45.5	862	2 T34342	hypothetical prote
9	45	45.5	892	2 T09071	SH3 domains-contai
10	45	45.5	1151	2 T30936	reverse transcript
11	44	44.4	282	2 A31765	heterogeneous ribo
12	44	44.4	296	2 AF2776	hypothetical prote
13	44	44.4	296	2 D97556	probable rmlb3 pro
14	44	44.4	376	2 C75580	adenine deaminase-
15	44	44.4	478	2 C87351	aldehyde dehydroge
16	43.5	43.9	486	2 T45702	hypothetical prote
17	43	43.4	142	2 E87334	hypothetical prote
18	43	43.4	203	2 C75366	chromosome partiti
19	43	43.4	257	2 D87689	dihydrodipicolinat
20	43	43.4	280	2 A69080	methylthioadenosin
21	43	43.4	407	2 D84169	molybdenum cofacto
22	43	43.4	585	2 T44309	H+-transporting tw
23	43	43.4	964	2 G71213	probable H(+)-tran
24	43	43.4	1017	2 D75028	H+-transporting AR
25	42.5	42.9	521	2 S54266	glycoprotein gC -
26	42.5	42.9	986	2 G65116	hypothetical prote
27	42.5	42.9	1128	1 T08312	plasmid replicatio
28	42	42.4	290	2 A26885	heterogeneous nucl
29	42	42.4	303	2 C34504	heterogeneous ribo

30	42	42.4	375	2 C72474	hypothetical prote
31	42	42.4	526	2 T29433	hypothetical prote
32	42	42.4	922	2 T23573	hypothetical prote
33	42	42.4	1367	2 T21913	hypothetical prote
34	42	42.4	1477	2 S64616	YOR1 protein - yea
35	42	42.4	2337	2 T40577	probable phosphati
36	41.5	41.9	693	1 E64139	DNA helicase recg
37	41.5	41.9	1266	2 A85989	hypothetical prote
38	41.5	41.9	1266	2 P91143	hypothetical prote
39	41	41.4	127	2 B91167	hypothetical prote
40	41	41.4	127	2 B86013	hypothetical prote
41	41	41.4	127	2 S47678	hypothetical 14.5K
42	41	41.4	206	2 JC2574	hypothetical 22K p
43	41	41.4	269	2 C82080	dihydrodipicolinat
44	41	41.4	322	1 C70714	probable epiA prot
45	41	41.4	328	2 A35663	CAMP response elem
46	41	41.4	419	2 T39248	probable GTP bindi
47	41	41.4	450	1 DCCHO	ornithine decarbox
48	41	41.4	477	2 H84222	hypothetical prote
49	41	41.4	483	2 F64760	membrane protein p
50	41	41.4	483	2 C90677	hypothetical prote
51	41	41.4	483	2 F85527	hypothetical prote
52	41	41.4	483	2 AI0547	PRPD protein limpo
53	41	41.4	1315	2 T05300	hypothetical prote
54	41	41.4	1319	2 S55598	tegument protein 0
55	41	41.4	1707	2 S77908	hypothetical prote
56	41	41.4	2051	2 T30938	receptor tyrosine
57	40	40.4	123	2 G72702	probable cytidine
58	40	40.4	171	2 T20567	hypothetical prote
59	40	40.4	185	2 S60778	dihydrodipicolinat
60	40	40.4	250	2 T35839	probable dihydrodi
61	40	40.4	263	2 T51322	nickel-insertion a
62	40	40.4	273	1 RDRCPD	dihydrodipicolinat
63	40	40.4	273	2 B99633	dihydrodipicolinat
64	40	40.4	273	2 A85484	dihydrodipicolinat
65	40	40.4	273	2 AC0510	dihydrodipicolinat
66	40	40.4	279	2 AF1155	hypothetical prote
67	40	40.4	281	2 T02813	hypothetical prote
68	40	40.4	304	2 C70978	probable rmlD prot
69	40	40.4	311	2 T43947	N-acetyl-gamma-glu
70	40	40.4	330	2 E75535	malate dehydrogena
71	40	40.4	335	2 A71042	probable mevalonat
72	40	40.4	408	2 JC5862	leucyl aminopeptid
73	40	40.4	417	2 AF0251	probable M23/M37 p
74	40	40.4	419	2 H64947	membrane protein y
75	40	40.4	419	2 B85798	hypothetical prote
76	40	40.4	419	2 F90949	membrane protein E
77	40	40.4	430	2 E82502	conserved hypothet
78	40	40.4	439	2 AF0742	conserved hypothet
79	40	40.4	439	2 T52291	probable DNA-bindi
80	40	40.4	455	2 T05081	hypothetical prote
81	40	40.4	488	2 A95926	probable argininos
82	40	40.4	490	2 G85354	hypothetical prote
83	40	40.4	503	2 AB2938	alpha-L-arabinofur
84	40	40.4	503	2 E98344	hypothetical prote
85	40	40.4	550	2 G70597	probable proteinas
86	40	40.4	585	2 JC5532	vacuolar-type ATPa
87	40	40.4	585	2 T44674	H+-transporting AR
88	40	40.4	623	2 T52293	MYC-related DNA bi
89	40	40.4	666	2 T10567	probable serine/th
90	40	40.4	790	2 D81064	hypothetical prote
91	40	40.4	957	2 F91100	glycine decarboxyl
92	40	40.4	957	2 B85946	hypothetical prote
93	40	40.4	957	2 S36834	glycine dehydrogen
94	40	40.4	1218	2 T30293	ABC transport prot
95	40	40.4	1218	2 AD0837	probable ABC trans
96	40	40.4	1549	2 T21809	hypothetical prote
97	40	40.4	2335	2 T40186	probable phosphati
98	39.5	39.9	186	2 E71335	probable cytidylat
99	39.5	39.9	317	2 T34843	probable transducti
100	39.5	39.9	462	2 G63220	sensory transducti

ALIGNMENTS

RESULT 1

D69512
 acetolactate synthase, large subunit (IlvB-4) homolog - Archaeoglobus fulgidus
 C:Species: Archaeoglobus fulgidus
 C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
 C:Accession: D69512
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
 .; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
 Nature 390, 364-370, 1997
 A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.
 Smith, H.O.; Woese, C.R.; Venter, J.C.
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archae
 A:Reference number: A69250; MUID:98049343; PMID:9389475
 A:Accession: D69512
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-575 <LE>
 A:Cross-references: GB:AE000958; GB:AE000782; NID:g2689281; PIDN:AB89146.1; PID:g264842
 C:Superfamily: acetolactate synthase large chain; thiamin pyrophosphate-binding domain h

Query Match 49.5%; Score 49; DB 2; Length 575;
 Best Local Similarity 62.5%; Pred. No. 9.1;
 Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 LQRAVRLALSGSDGRV 16

|||||

DB 147 VQRAIRALSGRPGV 162

RESULT 2

I48155
 Transcription initiation factor IID 250K chain splice form 2 - golden hamster
 N:Alternate names: 210K nuclear DNA-binding cell cycle gene 1 protein (CCG1); CCG1; TATA
 C:Species: Mesocricetus auratus (golden hamster)
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 27-Oct-2003
 C:Accession: I48155
 R:Hayashida, T.; Sekiguchi, T.; Noguchi, E.; Sunamoto, H.; Ohba, T.; Nishimoto, T.
 Gene 141, 267-270, 1994
 A:Title: The CCG1/TAFl1250 gene is mutated in thermosensitive G1 mutants of the BHK21 ce
 A:Reference number: I48155; MUID:94215915; PMID:8163200
 A:Accession: I48155
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1865 <RES>
 A:Cross-references: GB:D26114; NID:9439485; PIDN:BAA05110.1; PID:g474971
 C:Superfamily: TBP-associated factor 1 (TAF1), animal-type; bromodomain homology; HMG bo
 C:Keywords: alternative splicing; cell cycle control; DNA binding; duplication; phospho
 F:1190-1269/Domain: HMG box homology <HMG1>
 F:1346-1353/Region: nuclear location signal
 F:1400-1455/Domain: bromodomain homology <BRO1>
 F:1523-1578/Domain: bromodomain homology <BRO2>
 F:131.1714,1725,1821,1843/Binding site: phosphate (Ser) (covalent) (by casein kinase II)
 F:652,1028,1658/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kinase
 F:994,1335/Binding site: phosphate (Ser) (covalent) (by cAMP-dependent kinase) #status F
 F:1355,1374/Binding site: phosphate (Thr) (covalent) (by calmodulin-dependent kinase) #s

Query Match 48.5%; Score 48; DB 1; Length 1865;
 Best Local Similarity 62.5%; Pred. No. 42;
 Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 AVRLALSGSDGRVGS 19

|||||

DB 1735 AIQLSESGSDSDVGS 1750

RESULT 3

A40262
 Transcription initiation factor IID 250K chain splice form 1 - human
 N:Alternate names: 210K nuclear DNA-binding cell cycle gene 1 protein (CCG1); TATA-bindin

N:Contains: transcription initiation factor IID 250K chain splice form 2
 C:Species: Homo sapiens (man)
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 27-Oct-2003
 C:Accession: A40262; S03005; S00830; S32352; S32353
 R:Seiguchi, T.; Nohiro, Y.; Nakamura, Y.; Hisamoto, N.; Nishimoto, T.
 Mol. Cell. Biol. 11, 3317-3325, 1991
 A:Title: The human CCG1 gene, essential for progression of the G-1 phase, encodes a 210->

A:Reference number: A40262; MUID:91246200; PMID:2038334

A:Accession: A40262

A:Molecule type: mRNA

A:Residues: 1-177,199-1893 <SE>

A:Cross-references: GB:D90359; NID:g559319; PIDN:BAA14374.1; PID:g219528

A:Note: nucleotide sequence not complete

R:Seiguchi, T.; Miyata, T.; Nishimoto, T.

submitted to the EMBL Data Library, February 1988

A:Reference number: S03005

A:Accession: S03005

A:Molecule type: mRNA

A:Residues: 'MYR', 60-177,199-1604, 'DNECKSKANDIVCLIQCSSQIEELRF' <SES>

A:Cross-references: EMBL:X07024; NID:g29732; PIDN:CAA30073.1; PID:g29733

A:Note: this sequence has been revised in reference A40262

R:Seiguchi, T.; Miyata, T.; Nishimoto, T.

EMBO J. 7, 1683-1687, 1988

A:Title: Molecular cloning of the cDNA of human X chromosomal gene (CCG1) which compleme

A:Reference number: S00830; MUID:89005056; PMID:3169001

A:Accession: S00830

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1372-1379;1386-1604, 'DNE', 'CSSKANDIVCLIQCSSQIEELRF' <SE6>

A:Cross-references: EMBL:X07024

R:Ruppert, S.; Wang, E.H.; Tjian, R.

Nature 362, 175-179, 1993

A:Title: Cloning and expression of human TAF(II)250: a TBP-associated factor implicated i

A:Reference number: S32352; MUID:93196704; PMID:7680771

A:Accession: S32352

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 178-198 <RUP>

R:Hisatake, K.; Hasegawa, S.; Takada, R.; Nakatani, Y.; Horikoshi, M.; Roeder, R.G.

Nature 362, 179-181, 1993

A:Title: The p250 subunit of native TATA box-binding factor TFIID is the cell-cycle regul

A:Reference number: S32353; MUID:93196705; PMID:8450888

A:Accession: S32353

A:Status: preliminary

A:Molecule type: protein

A:Residues: 'P';587-595;1009-1022;1351-1355;1357-1360 <HIS>

A:Note: 1351-Val, 1353-Lys, and 1354-Glu were also found

C:Genetics:

A:Gene: GDB:TAF2A; CCG1; CCGS; NSCL2; TAFII250

A:Cross-references: GDB:120573; OMIM:313650

A:Map position: Xq13.1-Xq13.1

C:Superfamily: TBP-associated factor 1 (TAF1), animal-type; bromodomain homology; HMG bo

C:Keywords: alternative splicing; cell cycle control; DNA binding; duplication; phospho

F:1-1893/Product: transcription initiation factor IID 250K chain splice form 1 #status p

F:1-177,199-1893/Product: transcription initiation factor IID 250K chain splice form 2 #

F:1216-1295/Domain: HMG box homology <HMG1>

F:1372-1379/Region: nuclear location signal

F:1426-1481/Domain: bromodomain homology <BRO1>

F:1549-1604/Domain: bromodomain homology <BRO2>

F:137,1740,1751,1847,1871/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kinase)

F:678,1054,1684/Binding site: phosphate (Ser) (covalent) (by cAMP-dependent kinase) #status I

F:1020,1361/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kinase) #st

F:1381,1400/Binding site: phosphate (Thr) (covalent) (by calmodulin-dependent kinase) #st

Query Match 48.5%; Score 48; DB 1; Length 1893;

Best Local Similarity 62.5%; Pred. No. 42;

Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 AVRLALSGSDGRVGS 19

|||||

DB 1761 AIQLSESGSDSDVGS 1776

A:Protein sequence; Amino acid sequence not shown; Translation not shown
 A:Residues: 1-273 <MTH>
 A:Molecule type: DNA
 A:Cross-references: GB:AE000859; GB:AE000665; NID:g2621885; PIDN:AAB85300.1;
 A:Experimental source: strain Delta H
 C:GeneticCB:
 A:Gene: MTH800
 A:Start codon: TTG
 C:Superfamily: dihydrodipicolinate reductase

RESULT 7
E83312
hypothetical protein PA2671 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence revision 15-Sep-2000 #text change 31-Dec-2000

A:Accession: E83312
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-359 <STO>
 A:Cross-references: GB:AE004695; GB:AE004091; MID:G99948730; PIDN:AAG06059.1; GS
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: PA2671

hypothetical protein T08H4.1 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34342

A:Status: preliminary; translated from GE/EMBL/DBDJ
A:Accession: T08H4.1
A:Molecule type: DNA
A:Residues: 1-862 <STE>
A:Cross-references: EMBL:U39470, FIDN:AACT7118.1, GSPDB:GN00020; CESP:T08H4.1
A:Experimental source: strain Bristol N2; clone T08H4
C:Genetics:
A:Gene: CESP:T08H4.1
A:Map position: 2
A:Introns: 17/3; 44/2; 170/2; 259/3; 291/3; 553/1; 768/3; 804/3; 860/2

A:Map position: circular chromosome

Query Match 44.4%; Score 44; DB 2; Length 296;
Best Local Similarity 61.5%; Pred. No. 29;
Matches 8; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 5 VRLALSGSGRVG 17
DB 1 MRVLTGSGRVG 13

RESULT 14
C75580
adenine deaminase-related protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: C75580
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
S.; Smith, H.O.; Venter, J.C.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: C75580
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-376 <WHI>
A:Cross-references: GB:AE001863; GB:AE001825; NID:g6460670; PIDN:AAF12376.1; PID:g646067
A:Experimental source: strain R1
C:Genetics:
A:Gene: DRA0268
A:Map position: 2

Query Match 44.4%; Score 44; DB 2; Length 376;
Best Local Similarity 64.7%; Pred. No. 37;
Matches 11; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 3 RAVRALSGDGRVSG 19
DB 275 RAVAPALGSDDRPASG 291

RESULT 15
C87351
aldehyde dehydrogenase [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C:Accession: C87351
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: C87351
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-478 <STO>
A:Cross-references: GB:AE005673; NID:gl3422073; PIDN:AAK22807.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC0822
C:Superfamily: aldehyde dehydrogenase (NAD+); aldehyde dehydrogenase homology

Query Match 44.4%; Score 44; DB 2; Length 478;
Best Local Similarity 50.8%; Pred. No. 46;
Matches 10; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 LQRAVRLALSGDGRVSGC 20
DB 262 LERAVAILDGSFFATGQRC 281

RESULT 16

T45702
hypothetical protein F18L15.170 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T45702
R:Choisne, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Artiguenave, P.; Sa
submitted to the Protein Sequence Database, November 1999
A:Reference number: Z23010
A:Accession: T45702
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-486 <CHO>
A:Cross-references: EMBL:AL133298
A:Experimental source: cultivar Columbia; BAC clone F18L15
C:Genetics:
A:Map position: 3
A:Introns: 116/3; 169/2; 201/2; 287/1; 360/3
A:Note: F18L15.170

Query Match 43.9%; Score 43.5; DB 2; Length 486;
Best Local Similarity 45.8%; Pred. No. 56;
Matches 11; Conservative 4; Mismatches 4; Indels 5; Gaps 1;

QY 1 LQRAVRLALS-----GSDGRVSG 19
DB 32 IPRAVQQTISKRVSGSNGRFGSG 55

RESULT 17
E87334
hypothetical protein CC0688 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: E87334
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: E87334
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-142 <STO>
A:Cross-references: GB:AE005673; NID:gl3421909; PIDN:AAK22673.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC0688

Query Match 43.4%; Score 43; DB 2; Length 142;
Best Local Similarity 58.8%; Pred. No. 20;
Matches 10; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3 RAVRALSGDGRVSG 19
DB 95 RAVSLALDGDGRVYNG 111

RESULT 18
C75366
chromosome partitioning protein, ParA family - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: C75366
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
S.; Smith, H.O.; Venter, J.C.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: C75366
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-203 <WHI>

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A:Reference number: A71000; MUID:98344137; PMID:9679194
A:Accession: G71213
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-964 <KAW>
A:Cross-references: GB:AP000007; NID:G3236134; PIDN:BAA31102.1; PID:G3258419
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH1975
C:Superfamily: yeast vacuolar H⁺-transporting ATPase chain A; H⁺-transporting ATP synthase
F:633-815/Domain: H⁺-transporting ATP synthase alpha chain homology <ATP>
Query Match 43.4%; Score 43; DB 2; Length 964;
Best Local Similarity 58.8%; Pred. No. 1.3e+02;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 2 QRAVRLALSGSDGRVGS 18
DB 745 ERAGRVVTLGSDYRVGS 761
RESULT 24
D75028
H⁺-transporting ATP synthase, chain alpha (atpa) PAB2378 - Pyrococcus abyssi (strain Ors
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
A:Accession: D75028
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru
A:Reference number: A75001
A:Accession: D75028
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1017 <KAW>
A:Cross-references: GB:AJ248288; GB:AL096836; NID:G5458960; PIDN:CAB50666.1; PID:G545918
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB2378
C:Superfamily: yeast vacuolar H⁺-transporting ATPase chain A; H⁺-transporting ATP synthase
F:686-868/Domain: H⁺-transporting ATP synthase alpha chain homology <ATP>
Query Match 43.4%; Score 43; DB 2; Length 1017;
Best Local Similarity 58.8%; Pred. No. 1.4e+02;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 2 QRAVRLALSGSDGRVGS 18
DB 798 ERAGRVVTLGSDYRVGS 814
RESULT 25
S54266
glycoprotein gC - caprine herpesvirus 1
C:Species: caprine herpesvirus 1
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 26-Aug-1999
C:Accession: S54266
R:Recht, P.; Engels, M.; Loeppel, E.; Ackermann, M.
submitted to the EMBL Data Library, May 1995
A:Description: Comparison of the glycoprotein gC genes of bovine and caprine herpesvirus
A:Reference number: S54264
A:Accession: S54266
A:Molecule type: DNA
A:Residues: 1-521 <HEC>
A:Cross-references: EMBL:Z49225; NID:G804975; PIDN:CAA89200.1; PID:G804976
C:Superfamily: herpesvirus glycoprotein F
C:Keywords: glycoprotein
Query Match 42.9%; Score 42.5; DB 2; Length 521;
Best Local Similarity 61.1%; Pred. No. 87;
Matches 11; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

QY 3 RAVRLALSGSDGRVGS 20
DB 443 RSVRL-LSGADGPVATC 459

RESULT 26
G65116
hypothetical protein b3245 - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
A:Accession: G65116
R:Plattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Col
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: G65116
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-986 <BLAT>
A:Cross-references: GB:AE000403; GB:U00096; NID:G2367205; PIDN:AACT6277.1; PID:G2367206;
A:Experimental source: strain K-12, substrain MG1655
C:Superfamily: Escherichia coli hypothetical protein yhdP
Query Match 42.9%; Score 42.5; DB 2; Length 986;
Best Local Similarity 60.0%; Pred. No. 1.6e+02;
Matches 12; Conservative 1; Mismatches 4; Indels 3; Gaps 1;
QY 1 LQRAVRLALSGS---DGRVG 17
DB 477 LPEAVNEALSGSVADGKVG 496

RESULT 27
T08312
plasmid replication protein repI [similarity] - Halobacterium sp. (strain NRC-1) plasmid
N:Alternate names: hypothetical protein H1080
C:Species: Halobacterium sp.
A:Variety: strain NRC-1
C:Date: 18-Feb-2000 #sequence_revision 18-Feb-2000 #text_change 18-Feb-2000
C:Accession: T08312
R:Ng, W.V.; Ciuflo, S.A.; Smith, T.M.; Bumgarner, R.E.; Baskin, D.; Faust, J.; Hall, B.; I
Genome Res. 8, 1131-1141, 1998
A:Title: Snapshot of a large dynamic replicon in a halophilic Archaeon: megaplasamid or mi
A:Reference number: Z16408; MUID:99063795; PMID:9847077
A:Accession: T08312
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1128 <DAS>
A:Cross-references: EMBL:AF016485; NID:G2822278; PID:G2822373; HALOSP:H1080
A:Experimental source: strain NRC-1
C:Genetics:
A:Gene: repI; HALOSP:H1080
A:Genome: plasmid pNRC100
C:Superfamily: Halobacterium plasmid pNRC100 replication protein repH
C:Keywords: plasmid replication
Query Match 42.9%; Score 42.5; DB 1; Length 1128;
Best Local Similarity 57.9%; Pred. No. 1.9e+02;
Matches 11; Conservative 2; Mismatches 5; Indels 1; Gaps 1;
QY 1 LQRAVRLALSGSDGRVGS 19
DB 956 IQXWVR-ALLGSDGAVSRG 973

RESULT 28
A26885
heterogeneous nuclear ribonucleoprotein C - human
C:Species: Homo sapiens (man)
C:Date: 16-Aug-1988 #sequence_revision 16-Aug-1988 #text_change 24-Sep-1999
C:Accession: A26885

R;Swanson, M.S.; Nakagawa, T.Y.; LeVan, K.; Dreyfuss, G.
Mol. Cell. Biol. 7, 1731-1739, 1987
A:Title: Primary structure of human nuclear ribonucleoprotein particle C proteins: conserved structure
A:Reference number: A26885; MUID:8757872; PMID:3110598
A:Accession: A26885
A:Molecule type: mRNA
A:Residues: 1-290 <SWA>
A:Cross-references: GB:M16342; NID:g184266; PIDN:AAA52680.1; PID:g306875
C:Genetics:
A:Gene: GDB:SNRPC
A:Cross-references: GDB:118878
C:Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprotein repeat-containing proteins
C:Keywords: alternative splicing
F:17-77/Domain: ribonucleoprotein repeat homology <RRM1>

Query Match 42.4%; Score 42; DB 2; Length 290;
Best Local Similarity 44.4%; Pred. NO. 58;
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 2 QRAVRLALSGSDGRVGS 19
DB 60 ERNARAAGVAGEDGRMIAG 77

RESULT 29
C34504
heterogeneous ribonuclear particle protein C2 - human
C:Species: Homo sapiens (man)
C:Date: 22-Jun-1990 #sequence_revision 22-Jun-1990 #text_change 24-Sep-1999
C:Accession: C34504
R:Burd, C.G.; Swanson, M.S.; Goerlach, M.; Dreyfuss, G.
Proc. Natl. Acad. Sci. U.S.A. 86, 9788-9792, 1989
A:Title: Primary structures of the heterogeneous nuclear ribonucleoprotein A2, B1, and C proteins
A:Reference number: A34504; MUID:90099350; PMID:2557628
A:Accession: C34504
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-303 <BUR>
A:Cross-references: GB:M29063; NID:g337454; PIDN:AAA36576.1; PID:g337455
C:Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprotein repeat-containing proteins
C:Keywords: alternative splicing
F:17-77/Domain: ribonucleoprotein repeat homology <RRM1>

Query Match 42.4%; Score 42; DB 2; Length 303;
Best Local Similarity 44.4%; Pred. NO. 61;
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 2 QRAVRLALSGSDGRVGS 19
DB 60 ERNARAAGVAGEDGRMIAG 77

RESULT 30
C72474
hypothetical protein APE2436 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: C72474
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; KDNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix
A:Reference number: A72450; MUID:99310339; PMID:10382966
A:Accession: C72474
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-375 <KAW>
A:Cross-references: DDBJ:AP000064; NID:g5105945; PIDN:BAA81451.1; PID:d1045237; PID:g5105945
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE2436

Query Match 42.4%; Score 42; DB 2; Length 375;

Best Local Similarity 52.9%; Pred. NO. 75;
Matches 9; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 2 QRAVRLALSGSDGRVGS 18
DB 18 REAVCLVFDGSDGRLLS 34

Search completed: May 13, 2004, 06:53:21
Job time : 14 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2004, 06:49:47 ; Search time 7.66667 Seconds
(without alignments)
135.835 Million cell updates/sec

Title: 09549186-9

Perfect score: 99

Sequence: 1 LQRAVRLALSGDGRVSGC 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	48	48.5	1872	1	T2D1 HUMAN	P21675 homo sapien
2	47	47.5	273	1	DAPE_METH	O26891 methanobact
3	46	46.5	393	1	DCO2 DROME	P40808 drosophila
4	45	45.5	394	1	DCO1 DROME	P40807 drosophila
5	45	45.5	592	1	VATA_METKA	Q8twl6 methanopyru
6	44	44.4	282	1	ROC_XENLA	P19600 xenopus lae
7	43	43.4	257	1	DAPE_CAUCR	Q9a2l1 caulobacter
8	43	43.4	585	1	VATA_THESI	O32466 thermococcu
9	43	43.4	964	1	VATA_PYRHO	O57728 pyrococcus
10	43	43.4	1013	1	VATA_PYRFU	Q8u4a6 pyrococcus
11	43	43.4	1017	1	VATA_PYRAB	Q9uxu7 pyrococcus
12	43	43.4	3093	1	POLG_BSTV1	Q65730 b genome po
13	42.5	42.9	986	1	YHDP_ECOLI	P46474 escherichia
14	42	42.4	306	1	ROC_HUMAN	P07910 homo sapien
15	42	42.4	313	1	ROC_MOUSE	Q92204 mus musculu
16	42	42.4	546	1	DX56 MOUSE	Q9d0r4 mus musculu
17	42	42.4	547	1	DX56 HUMAN	Q9ny93 homo sapien
18	42	42.4	593	1	VATA_PYRAE	Q8zvr1 pyrobaculum
19	42	42.4	1477	1	YORI_YEAST	P53049 saccharomyc
20	42	42.4	2337	1	TOR2_SCHPO	Q9y7k2 schizosacch
21	41.5	41.9	693	1	RECQ_HAEIN	P34809 haemophilus
22	41	41.4	127	1	YHKK_ECOLI	P37613 escherichia
23	41	41.4	146	1	CDD_MOUSE	P56389 mus musculu
24	41	41.4	269	1	DAPE_VIBCH	Q9kph7 vibrio chol
25	41	41.4	269	1	DAPE_VIBPA	O87sf5 vibrio para
26	41	41.4	450	1	DCOR_CHICK	P27118 gallus gall
27	41	41.4	482	1	PRPD_ECO57	Q8x693 escherichia
28	41	41.4	482	1	PRPD_ECOLI	P77243 escherichia
29	41	41.4	483	1	PRPD_SALTI	Q82903 salmonella
30	41	41.4	483	1	PRPD_SALTY	P74840 salmonella
31	41	41.4	677	1	RGSR_RAT	P49805 rattus norv
32	40	40.4	185	1	DAPE_KLEPN	P45415 klebsiella
33	40	40.4	250	1	DAPE_STRCO	O86836 streptomyce

34	40	40.4	258	1	DAB2_RHILO	P58211 rhizobium l
35	40	40.4	263	1	COOC_RHORU	P31897 rhodospiril
36	40	40.4	269	1	DABP_VIBVU	Q8d6m0 vibrio vuln
37	40	40.4	269	1	DABP_VIBVY	Q7muu2 vibrio vuln
38	40	40.4	273	1	DABP_ECO57	P58209 escherichia
39	40	40.4	273	1	DABP_ECOL6	Q8f1b4 escherichia
40	40	40.4	273	1	DABP_ECOLI	P04036 escherichia
41	40	40.4	273	1	DABP_SALTI	Q82919 salmonella
42	40	40.4	273	1	DABP_SALTY	Q829x8 salmonella
43	40	40.4	273	1	DABP_SHIFL	Q83sg9 shigella fl
44	40	40.4	335	1	KIME_PVRHO	Q59291 pyrococcus
45	40	40.4	344	1	ARC2_THETH	O52146 thermus the
46	40	40.4	396	1	EFTU_THICU	P42481 thioacillu
47	40	40.4	408	1	AMPT_THETH	P24778 thermus the
48	40	40.4	440	1	YEBA_ECOLI	P24204 escherichia
49	40	40.4	488	1	ARL2_RHIME	Q92ym6 rhizobium m
50	40	40.4	570	1	TREA_SALTI	P59765 salmonella
51	40	40.4	570	1	TREA_SALTY	Q82920 salmonella
52	40	40.4	585	1	VATA_DESSY	O06504 desulfuroco
53	40	40.4	623	1	RAP1_ARATH	Q39204 arabidopsis
54	40	40.4	728	1	CATB_AJECA	Q9y7c2 ajellomyces
55	40	40.4	956	1	GCSP_ECO57	Q8xd33 escherichia
56	40	40.4	956	1	GCSP_ECOL6	Q8f6e7 escherichia
57	40	40.4	956	1	GCSP_ECOLI	P33195 escherichia
58	40	40.4	1549	1	SMC4_CABEL	Q20960 caenorhabdi
59	40	40.4	2335	1	TOR1_SCHPO	O14356 schizosacch
60	39.5	39.9	471	1	YEGB_ECOLI	P36554 escherichia
61	39.5	39.9	493	1	VL2_PAPVD	P03110 deer papill
62	39	39.4	17	1	MDH_ACIDE	P08540 acidovorax
63	39	39.4	151	1	GDF7_MOUSE	P43029 mus musculu
64	39	39.4	222	1	DEOC_ECOLI	Q8nc44 corynebacte
65	39	39.4	229	1	MODB_ECOLI	P09834 escherichia
66	39	39.4	257	1	DABP_ARCFU	O29353 archaeoglob
67	39	39.4	269	1	DABP_BORBR	Q7wq55 bordetella
68	39	39.4	269	1	DABP_BORPA	Q7w510 bordetella
69	39	39.4	269	1	DABP_BORPE	Q9x5y9 bordetella
70	39	39.4	313	1	HXAB_MOUSE	P31311 mus musculu
71	39	39.4	328	1	YJ77_YEAST	P47145 saccharomyc
72	39	39.4	332	1	HBPB_WHEAT	P23923 triticum ae
73	39	39.4	355	1	MOAA_AERPE	Q9yev3 aeropyrum p
74	39	39.4	400	1	NRB_AZOB	P45670 azospirillu
75	39	39.4	431	1	DCOR_DATST	P50134 datura atra
76	39	39.4	484	1	XYLB_ECOLI	P09099 escherichia
77	39	39.4	564	1	OXC_ECOLI	P78093 escherichia
78	39	39.4	568	1	OXC_OXAFO	P40149 oxalobacter
79	39	39.4	782	1	FIBA_RAT	P06399 rattus norv
80	39	39.4	791	1	KDGL_DROME	Q01583 drosophila
81	39	39.4	1019	1	TOP1_USTMA	P41311 ustilago ma
82	39	39.4	1691	1	POLN_HEVME	Q03495 hepatitis e
83	38.5	38.9	458	1	ME31_DROME	P23128 drosophila
84	38.5	38.9	598	1	RACA_DICDI	P34147 dictyosteli
85	38	38.4	22	1	MDH_XANMA	P06541 xanthomonas
86	38	38.4	111	1	LV6D_HUMAN	P06318 homo sapien
87	38	38.4	240	1	COMB_DEIRA	Q9ru16 deinococcus
88	38	38.4	265	1	DABP_WIGBR	Q8d3h6 wigglewort
89	38	38.4	266	1	THID_RHIME	P56904 rhizobium m
90	38	38.4	269	1	DABP_NEIMA	Q9jx48 neisseria m
91	38	38.4	269	1	DABP_NEIMA	Q9k1f1 neisseria m
92	38	38.4	270	1	DABP_SHEON	Q8eb57 shewanella
93	38	38.4	273	1	DABP_BUCAP	Q8k325 buchnera ap
94	38	38.4	273	1	DABP_YERPE	Q82116 yersinia pe
95	38	38.4	293	1	NANA_STAAM	Q99wr1 staphylococ
96	38	38.4	293	1	NANA_STAAW	Q8nvc7 staphylococ
97	38	38.4	318	1	PTA_PARDE	P31917 paracoccus
98	38	38.4	323	1	FASH_AGR5	Q8ue62 agrobacteri
99	38	38.4	327	1	MDH_THETH	P10584 thermus the
100	38	38.4	343	1	RADA_HALVO	Q48328 halobacteri

ALIGNMENTS

T2D1 HUMAN
 ID T2D1 HUMAN STANDARD; PRT; 1872 AA.
 AC P21675;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Transcription initiation factor TFIID 250 kDa subunit (TAFII-250)
 DE (TAFII250) (TBP-associated factor 250 kDa) (P250) (Cell cycle gene 1 protein)
 DE Protein
 GN TAFI OR TAF2A OR CCG1 OR BAZR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Laryngeal carcinoma;
 RX MEDLINE=91246200; PubMed=2038334;
 RA Sekiguchi T., Nohiro Y., Nakamura Y., Hisamoto N., Nishimoto T.;
 RT "The human CCG1 gene, essential for progression of the G1 phase, encodes a 210-kilodalton nuclear DNA-binding protein.";
 RT Mol. Cell. Biol. 11:3317-3325(1991).
 RL [2]
 RN [3]
 RP PRELIMINARY SEQUENCE FROM N.A.
 RX MEDLINE=89005056; PubMed=3169001;
 RA Sekiguchi T., Miyata T., Nishimoto T.;
 RT "Molecular cloning of the cDNA of human X chromosomal gene (CCG1) which complements the temperature-sensitive G1 mutants, tsBN462 and ts13, of the BHK cell line.";
 RT EMBO J. 7:1683-1687(1988).
 RL [3]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RX MEDLINE=93196704; PubMed=7680771;
 RA Ruppert S., Wang E.H., Tjian R.;
 RT "Cloning and expression of human TAFII250: a TBP-associated factor implicated in cell-cycle regulation.";
 RL Nature 362:175-179(1993).
 RN [4]
 RP CHARACTERIZATION.
 RX MEDLINE=93196705; PubMed=8450888;
 RA Hisatake K., Hasegawa S., Takada R., Nakatani Y., Horikoshi M., Roeder R.G.;
 RT "The p250 subunit of native TATA box-binding factor TFIID is the cell-cycle regulatory protein CCG1.";
 RL Nature 362:179-181(1993).
 CC -1- FUNCTION: May play an essential role in TFIID assembly by interacting with both TBP and other TAF, as well as serving to link the control of transcription to the cell cycle. Essential for progression of the G1 phase of the cell cycle. Possesses DNA-binding activity.
 CC -1- SUBUNIT: TFIID is composed of TBP and a variety of TBP-associated factors.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- PTM: Phosphorylated by casein kinase II in vitro.
 CC -1- SIMILARITY: Contains 2 bromodomains.
 CC -1- SIMILARITY: Contains 1 HMG box domain.
 CC -1- SIMILARITY: TO DROSOPHILA TAFII-230. SOME TO S.POMBE TAFII-111 AND TO S.CEREVISIAE TAF145.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D90359; BA014374.1; -;
 CC EMBL; X07024; CAA30073.1; ALT_SEQ.
 CC F01; A0262; A40262.
 CC F01; 1EQF; 07-JUN-00.
 CC TRANSFAC; T02206; -;
 CC Genew; HGNC:11535; TAF1.

MIM; 313650; -;
 DR GO; GO:0005669; C:transcription factor TFIID complex; TAS.
 DR GO; GO:0016251; F:general RNA polymerase II transcription fac. . .; TAS.
 DR GO; GO:0004672; F:protein kinase activity; TAS.
 DR GO; GO:0000114; P:G1-specific transcription in mitotic cell c. . .; TAS.
 DR GO; GO:0006468; P:protein amino acid phosphorylation; TAS.
 DR GO; GO:0006461; P:protein complex assembly; TAS.
 DR InterPro; IPR001487; Bromodomain.
 DR Pfam; PF00439; bromodomain; 2.
 DR PRINTS; PRO0503; bromodomain.
 DR SMART; SM00297; BROMO; 2.
 DR PROSITE; PS00633; BROMODOMAIN 1; 2.
 DR PROSITE; PS0014; BROMODOMAIN 2; 2.
 KW Bromodomain; Nuclear protein; DNA-binding; Cell cycle; Repeat;
 KW Transcription regulation; Phosphorylation; 3D-structure.
 FT DOMAIN 157 165 PRO-RICH.
 FT DNA BIND 1195 1273 HMG BOX (POTENTIAL).
 FT DOMAIN 1351 1358 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 1397 1467 BROMODOMAIN 1.
 FT DOMAIN 1520 1590 BROMODOMAIN 2.
 FT DOMAIN 1627 1872 ASP/GLU-RICH (ACIDIC TAIL).
 SQ SEQUENCE 1872 AA; 212676 MW; 93BE3D181A72ABEB CRC64;
 Query Match 48.5%; Score 48; DB 1; Length 1872;
 Best Local Similarity 62.5%; Pred. No. 23;
 Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 4 AVRLALSSGSDGRVSG 19
 DB 1740 AIQLSESGSDSDVSG 1755
 RESULT 2
 DABP METH
 ID DABP METH STANDARD; PRT; 273 AA.
 AC Q26891;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Dihydrodipicolinate reductase (EC 1.3.1.26) (DHPR).
 GN DABP OR MTH800.
 OS Methanobacterium thermoautotrophicum.
 OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
 OC Methanobacteriaceae; Methanothermobacter.
 OX NCBI_TaxID=187420;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Delta H;
 RX MEDLINE=98037514; PubMed=9371463;
 RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J., Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K., Harrison D., Hoang L., Reagle P., Lum W., Pothier B., Qiu D., Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R., Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S., McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M., Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
 RT "Complete genome sequence of Methanobacterium thermoautotrophicum deltaH: functional analysis and comparative genomics.";
 RL J. Bacteriol. 179:7135-7155(1997).
 CC -1- CATALYTIC ACTIVITY: 2,3,4,5-tetrahydrodipicolinate + NAD(P) (+) = 2,3-dihydrodipicolinate + NAD(P)H.
 CC -1- PATHWAY: Biosynthesis of diaminopimelate and lysine from aspartate semialdehyde; second step.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: belongs to the dihydrodipicolinate reductase family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```
CC -----
DR EMBL; AF000858; AAB85300.1; -.
DR PIR; H69206; H69206.
DR HSSP; P04036; 1DRW.
DR HAMAP; MF_00102; -. 1.
DR InterPro; IPR000846; DapB.
DR Pfam; PF05173; DapB_C; 1.
DR Pfam; PF01113; DapB_N; 1.
DR ProDom; PD004105; DapB; 1.
DR TIGRFAMs; TIGR00036; dapB; 1.
DR PROSITE; PS01298; DAPB; 1.
DR Diaminopimelate biosynthesis; Lysine biosynthesis; Oxidoreductase;
KW NADP; Complete proteome.
SQ SEQUENCE 273 AA; 29045 MW; 5242C2A5D7B59B15 CRC64;

Query Match 47.5%; Score 47; DB 1; Length 273;
Best Local Similarity 53.3%; Pred. No. 4;
Matches 8; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 5 VRLALSGSDGRVSGS 19
   :|:|:|:|:|:|
Db 2 IRVAVTGACGRMGSG 16

RESULT 3
DCO2 DROME
ID DCO2 DROME STANDARD; PRT; 393 AA.
AC P40808;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ornithine decarboxylase 2 (EC 4.1.1.17) (ODC).
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93319633; PubMed=8329117;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA Sutton G.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.P., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balieu R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Galbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hosin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mileshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weissstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
```

```
FT BINDING 62 62 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT ACT SITE 343 343 BY SIMILARITY.
SQ SEQUENCE 393 AA; 44140 MW; A6F56F94CIAD5836 CRC64;

Query Match 46.5%; Score 46; DB 1; Length 393;
Best Local Similarity 45.0%; Pred. No. 8.6;
Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 1 LQRAVRLALSGSDGRVSGS 20
   :|:|:|:|:|
Db 177 LAKALGLKVTGTSFHVSGS 196

RESULT 4
DCO1 DROME
ID DCO1 DROME STANDARD; PRT; 394 AA.
AC P40807; Q9V352;
DT 01-FEB-1995 (Rel. 31, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ornithine decarboxylase 1 (EC 4.1.1.17) (ODC).
GN ODC1 OR CG8721.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Canton-S;
RC MEDLINE=93319633; PubMed=8329117;
RA Rom E., Kahana C.;
RT "Isolation and characterization of the Drosophila ornithine
RT decarboxylase locus: evidence for the presence of two transcribed ODC
RT genes in the Drosophila genome.";
RL DNA Cell Biol. 12:499-508(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA Sutton G.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.P., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balieu R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Galbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hosin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mileshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weissstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
```

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
CC -|- CATALYTIC ACTIVITY: L-ornithine = putrescine + CO(2).
CC -|- COFACTOR: Pyridoxal phosphate.
CC -|- PATHWAY: Polyamine biosynthesis; first (rate-limiting) step.
CC -|- SIMILARITY: BELONGS TO FAMILY 2 OF ORNITHINE, DAP, AND ARGinine
CC DECARBOXYLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X66601; CAA47167.1; -.
DR EMBL; X66599; CAA47165.1; -.
DR EMBL; AE003839; AAF59150.1; -.
DR HSSP; P00860; 70DC.
DR FlyBase; FBgn0013307; Odc1.
DR InterPro; IPR000183; Decarboxylase2.
DR InterPro; IPR009006; Racem decarbox_C.
DR Pfam; PF02784; Orn_Arg_dec_N; 1.
DR Pfam; PF00278; Orn_DAP_Arg_dec; 1.
DR PRINTS; PR01179; ODACRBLKASE.
DR PROSITE; PS00878; ODR_DC_2.1; 1.
DR PROSITE; PS00879; ODR_DC_2.2; 1.
DR Lyase; Decarboxylase; Pyridoxal phosphate; Polyamine biosynthesis;
KW Multigene family.
KW BINDING 62 62 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
FT ACT_SITE 343 343 BY SIMILARITY.
FT CONFLICT 44 45 NV -> KL (IN REF. 2).
FT CONFLICT 242 242 K -> Q (IN REF. 1; CAA47165).
FT CONFLICT 325 325 L -> Q (IN REF. 1; CAA47165).
FT CONFLICT 348 348 K -> Q (IN REF. 1; CAA47165).
SQ SEQUENCE 394 AA; 44166 MW; CD90520632B7D3B2 CRC64;
Query Match 45.5%; Score 45; DB 1; Length 394;
Best Local Similarity 40.0%; Pred. No. 12;
Matches 8; Conservative 6; Mismatches 6; Indels 0; Gaps 0;
QY 1 LQAVRLALSGSDGRVSGC 20
DB 177 LAKSLKVTGTGFHVGSGC 196
RESULT 5
VATA_METKA
ID VATA_METKA STANDARD; PRT; 592 AA.
AC Q8TWL6;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE V-type ATP synthase alpha chain (EC 3.6.3.14) (V-type ATPase subunit
DE A).
GN ATPA OR MK1017.
OS Methanopyrus kandleri.
OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;
OC Methanopyrus.
OX NCBI_TaxID=2320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AV19 / DSM 6324 / JCM 9639;
RX MEDLINE=21927647; PubMed=11930014;
RA Slesarev A.I., Mezheva K.V., Makarova K.S., Polushin N.N.,
RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,
RA Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,
RA Malykh A.G., Koonin E.V., Kozlovskiy S.A.;
RT "The complete genome of hyperthermophile *Methanopyrus kandleri* AV19

RT and monophyly of archaeal methanogens.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).
CC -|- FUNCTION: Produces ATP from ADP in the presence of a proton
CC gradient across the membrane. The archaeal alpha chain is a
CC catalytic subunit.
CC -|- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
CC H(+) (Out).
CC -|- SIMILARITY: Belongs to the ATPase alpha/beta chains family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE010390; AAM02230.1; -.
DR HAMAP; MF 00309; -; 1.
DR InterPro; IPR003593; AAA ATPase.
DR InterPro; IPR005726; ATP_syntha_A.
DR InterPro; IPR000793; ATPase_a/b_C.
DR InterPro; IPR000194; ATPase_a/bcentre.
DR InterPro; IPR004100; ATPase_a/bN.
DR InterPro; IPR009005; F1 ATPase_a/bN.
DR Pfam; PF00006; ATP-synt_ab; 1.
DR Pfam; PF00306; ATP-synt_ab_C; 1.
DR Pfam; PF02874; ATP-synt_ab_N; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRfam; TIGR01043; ATP_syn_A_arch; 1.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
KW Hydrolase; ATP synthetase; Hydrogen ion transport; ATP-binding;
KW Complete proteome. 243 ATP (POTENTIAL).
FT NP_BIND 236 243
SQ SEQUENCE 592 AA; 66494 MW; 3E0E1B54B395B6E1 CRC64;
Query Match 45.5%; Score 45; DB 1; Length 592;
Best Local Similarity 58.8%; Pred. No. 19;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 2 QRAVRLALSGSDGRVGS 18
DB 371 ERAGRVVCLGSDRVS 387
RESULT 6
ROC_XENLA
ID ROC_XENLA STANDARD; PRT; 282 AA.
AC P19600;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Heterogeneous nuclear ribonucleoprotein C (hnRNP core protein C).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89071757; PubMed=2904678;
RA Preugschat F., Wold B.;
RT "Isolation and characterization of a Xenopus laevis C protein cDNA:
RT structure and expression of a heterogeneous nuclear ribonucleoprotein
RT core protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:9669-9673(1988).
CC -|- FUNCTION: May play a role in nucleosome assembly by neutralizing
CC basic proteins such as A and B core hnRNPs.
CC -|- SUBCELLULAR LOCATION: Nuclear; component of ribonucleosomes.
CC -|- PTM: Phosphorylated (Probable).
CC -|- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; J03831; AAA60937.1; -.
 DR PIR; A31765; A31765.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR Pfam; PF00076; rrm; 1.
 DR SMART; SM00360; RRM; 1.
 DR PROSITE; PS0102; RRM; 1.
 DR PROSITE; PS00030; RRM RNP 1; 1.
 KW Nuclear protein; RNA-binding; Ribonucleoprotein; Phosphorylation.
 FT DOMAIN 17 88 RNA-BINDING (RRM).
 FT DOMAIN 141 147 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 178 282 ASP/GLU-RICH (ACIDIC).
 FT MOD_RES 240 240 PHOSPHORYLATION (POTENTIAL).
 FT MOD_RES 256 256 PHOSPHORYLATION (POTENTIAL).
 FT MOD_RES 267 267 PHOSPHORYLATION (POTENTIAL).
 SQ SEQUENCE 282 AA; 30950 MW; 7373FA46F8C85413 CRC64;

Query Match 44.4%; Score 44; DB 1; Length 282;
 Best Local Similarity 44.4%; Pred. No. 13;
 Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 2 QRAVRLALSGSDGRVGS 19
 DB 61 ERTARTAVAGEDGRMIAG 78

RESULT 7

DABP CAUCR
 ID DABP CAUCR STANDARD; PRT; 257 AA.
 AC Q9AZL1;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Dihydropicolinate reductase (EC 1.3.1.26) (DHPR).
 GN DABP OR CC3550.
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;
 OC Caulobacteraceae; Caulobacter.
 OX NCBI_TaxID=155892;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 19089 / CB15;
 RX MEDLINE=21173698; PubMed=11259647;
 RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
 RA Eisen J.A., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
 RA Kolonay J.P., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
 RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
 RT "Complete genome sequence of *Caulobacter crescentus*.";
 Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 CC -1- CATALYTIC ACTIVITY: 2,3,4,5-tetrahydropicolinate + NAD(P)(+) =
 CC 2,3-dihydropicolinate + NAD(P)H.
 CC -1- PATHWAY: Biosynthesis of diaminopimelate and lysine from aspartate
 CC semialdehyde; second step.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: Belongs to the dihydropicolinate reductase family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AE006013; AAK25512.1; -.
 DR PIR; D87689; D87689.
 DR HSSP; P04036; 1DRW.
 DR TIGR; CC3550; -.
 DR HAMAP; MF_00102; -. 1.
 DR InterPro; IPR000846; DapB.
 DR Pfam; PF05173; DapB_C; 1.
 DR Pfam; PF01113; DapB_N; 1.
 DR ProDom; PD004105; DapB; 1.
 DR TIGRFAMs; TIGR00036; dapB; 1.
 DR PROSITE; PS01298; DAPB; 1.
 KW Diaminopimelate biosynthesis; Lysine biosynthesis; Oxidoreductase;
 KW NADP; Complete proteome.
 SQ SEQUENCE 257 AA; 26425 MW; B8AD3B269A47AE4 CRC64;

Query Match 43.4%; Score 43; DB 1; Length 257;
 Best Local Similarity 35.3%; Pred. No. 16;
 Matches 6; Conservative 9; Mismatches 2; Indels 0; Gaps 0;

QY 1 LQRAVRLALSGSDGRVG 17
 DB 1 MSQPVKIALAGANGRMG 17

RESULT 8

VATA THESI
 ID VATA THESI STANDARD; PRT; 585 AA.
 AC Q32466;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE V-type ATP synthase alpha chain (EC 3.6.3.14) (V-type ATPase subunit
 DE A).
 GN ATPA.
 OS Thermococcus sp. (strain KI).
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
 OC Thermococcus.
 OX NCBI_TaxID=35749;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98036007; PubMed=9370240;
 RA Iida T., Kanai S., Inatomi K., Kanagata Y., Maruyama T.;
 RT "Alpha- and beta-subunits of a V-type membrane ATPase in a
 RT hyperthermophilic sulfur-dependent archaeum, *Thermococcus* sp. KI.";
 RL Biochim. Biophys. Acta 1329:12-17(1997).
 CC -1- FUNCTION: Produces ATP from ADP in the presence of a proton
 CC gradient across the membrane. The archaeal alpha chain is a
 CC catalytic subunit.
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
 CC H(+) (Out).
 CC -1- SIMILARITY: Belongs to the ATPase alpha/beta chains family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; D88772; BAA23342.1; -.
 DR PIR; T44309; T44309.
 DR HAMAP; MF_00309; -. 1.
 DR InterPro; IPR003593; AAA ATPase.
 DR InterPro; IPR005726; ATP_syntha_A.
 DR InterPro; IPR000793; ATPase_a/b_C.
 DR InterPro; IPR000194; ATPase_a/bcentre.
 DR InterPro; IPR004100; ATPase_a/bN.
 DR InterPro; IPR009005; F1 ATPase_a/bN.
 DR Pfam; PF00006; ATP-synt_ab; 1.
 DR Pfam; PF00306; ATP-synt_ab_C; 1.
 DR Pfam; PF02874; ATP-synt_ab_N; 1.
 DR SMART; SM00382; AAA; 1.

```

DR TIGRFAMs; TIGR01043; ATP syn A arch; 1.
DR PROSITE; PS00152; ATPase ALPHA_BETA; 1.
KW Hydrolase; ATP synthesis; Hydrogen ion transport; ATP-binding.
FT NP_BIND 231 238 ATP (POTENTIAL).
SQ SEQUENCE 585 AA; 65485 MW; B1653C9AC2F25054 CRC64;

Query Match 43.4%; Score 43; DB 1; Length 585;
Best Local Similarity 58.8%; Pred. No. 40;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 QRAVRLALSGSDGRVGS 18
Db 366 ERAGRVTILGSDRVGS 382

RESULT 9
VATA PYRHO
ID VATA PYRHO STANDARD; PRT; 964 AA.
AC O57728;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE V-type ATP synthase alpha chain (EC 3.6.3.14) (V-type ATPase subunit
DE A) [Contains: Endonuclease PI-Pho2 (EC 3.1.1.-) (Pho atpA intein) (Pho
DE VMA intein)].
GN ATPA OR PH1975.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
CC -!- FUNCTION: Produces ATP from ADP in the presence of a proton
CC gradient across the membrane. The archaeal alpha chain is a
CC catalytic subunit.
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
CC H(+) (Out).
CC -!- PTM: This protein undergoes a protein self splicing that involves
CC a post-translational excision of the VDE intervening region
CC (intein) followed by peptide ligation (Potential).
CC -!- MISCELLANEOUS: The intein interrupts the ATP-binding site.
CC -!- SIMILARITY: Belongs to the ATPase alpha/beta chains family.
CC -!- SIMILARITY: In the intein section; belongs to the homing
CC endonuclease family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF000007; BAA31102.1; -.
CC PIR; G71213.
CC HAMAP; MF_00309; atypical; 1.
DR InterPro; IPR000793; ATPase_a/b_C.
DR InterPro; IPR000194; ATPase_a/bcentre.
DR InterPro; IPR004100; ATPase_a/bN.
DR InterPro; IPR009005; Fl_ATPase_a/bN.
DR InterPro; IPR003587; Hedgehog hint N.
DR InterPro; IPR003586; Hedgehog hintC.

```

```

DR InterPro; IPR004042; Intein endonuc.
DR InterPro; IPR006141; Intein S.
DR Pfam; PF00006; ATP-synt_ab_1.
DR Pfam; PF00306; ATP-synt_ab_C; 1.
DR Pfam; PF02874; ATP-synt_ab_N; 1.
DR SMART; SM00305; HintC; 1.
DR SMART; SM00306; HintN; 1.
DR TIGRFAMs; TIGR01443; intein Cterm; 1.
DR TIGRFAMs; TIGR01445; intein Nterm; 1.
DR PROSITE; PS00152; ATPase ALPHA_BETA; 1.
DR PROSITE; PS00818; INTEIN_C_TER; 1.
DR PROSITE; PS00819; INTEIN_N_TER; 1.
DR PROSITE; PS00817; INTEIN_N_TER; 1.
KW Hydrolase; ATP synthesis; Hydrogen ion transport; ATP-binding;
KW Autocatalytic cleavage; Protein splicing; Nuclease; Endonuclease;
KW Intron homing; Complete proteome.
FT CHAIN 1 240 V-TYPE ATP SYNTHASE ALPHA CHAIN, 1ST
FT PART (POTENTIAL).
FT CHAIN 241 616 PHO ATPA INTEIN (POTENTIAL).
FT CHAIN 617 964 V-TYPE ATP SYNTHASE ALPHA CHAIN, 2ND
FT PART (POTENTIAL).
SQ SEQUENCE 964 AA; 107854 MW; 33252C47713BD5E1 CRC64;

Query Match 43.4%; Score 43; DB 1; Length 964;
Best Local Similarity 58.8%; Pred. No. 69;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 QRAVRLALSGSDGRVGS 18
Db 745 ERAGRVTILGSDRVGS 761

RESULT 10
VATA PYRPU
ID VATA PYRPU STANDARD; PRT; 1013 AA.
AC Q8U4A6;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE V-type ATP synthase alpha chain (EC 3.6.3.14) (V-type ATPase subunit
DE A) [Contains: Endonuclease PI-Pfu2 (EC 3.1.1.-) (Pfu atpA intein) (Pfu
DE VMA intein)].
GN ATPA OR PF0182.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Produces ATP from ADP in the presence of a proton
CC gradient across the membrane. The archaeal alpha chain is a
CC catalytic subunit.
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
CC H(+) (Out).
CC -!- PTM: This protein undergoes a protein self splicing that involves
CC a post-translational excision of the VDE intervening region
CC (intein) followed by peptide ligation (Potential).
CC -!- MISCELLANEOUS: The intein interrupts the ATP-binding site.
CC -!- SIMILARITY: Belongs to the ATPase alpha/beta chains family.
CC -!- SIMILARITY: In the intein section; belongs to the homing
CC endonuclease family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

CC EMBL; A5010144; AAL80306.1; -.
DR HAMAP; MF_00309; atypical; 1.
DR InterPro; IPR000793; ATPase_a/b C.
DR InterPro; IPR000194; ATPase_a/bcentre.
DR InterPro; IPR0004100; ATPase_a/bN.
DR InterPro; IPR009005; F1_ATPase_a/bN.
DR InterPro; IPR003587; Hedgehog_hint_N.
DR InterPro; IPR003586; Hedgehog_hintC.
DR InterPro; IPR006142; INTEIN.
DR InterPro; IPR004042; Intein_endonuc.
DR InterPro; IPR006141; Intein_S.
DR Pfam; PF00006; ATP-synt_ab; 1.
DR Pfam; PF00306; ATP-synt_ab_C; 1.
DR Pfam; PF02874; ATP-synt_ab_N; 1.
DR PRINTS; PR00379; INTEIN.
DR SMART; SM00305; HintC; 1.
DR SMART; SM00306; HintN; 1.
DR TIGRFAMs; TIGR01443; intein_Cterm; 1.
DR TIGRFAMs; TIGR01445; intein_Nterm; 1.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
DR PROSITE; PS50818; INTEIN_C_TER; 1.
DR PROSITE; PS50819; INTEIN_N_TER; 1.
DR PROSITE; PS50817; INTEIN_ENDONUCLEASE; 1.
DR Hydroxylase; ATP synthase; Hydrogen ion transport; ATP-binding;
KW Autocatalytic cleavage; Nuclease; Endonuclease;
KW Intronic homing; Complete proteome.
FT CHAIN 1 240 V-TYPE ATP SYNTHASE ALPHA CHAIN, 1ST
FT CHAIN 241 665 PFU ATPA INTEIN (POTENTIAL).
FT CHAIN 666 1013 V-TYPE ATP SYNTHASE ALPHA CHAIN, 2ND
FT CHAIN PART (POTENTIAL).
SQ SEQUENCE 1013 AA; 113452 MW; CAC1CB86566C81CD CRC64;

Query Match 43.4%; Score 43; DB 1; Length 1013;
Best Local Similarity 58.8%; Pred. No. 72;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 QRAVRLALSGSDGRVGS 18
Db :|||: |||||
794 ERAGRVTGLSDYRVGS 810

RESULT 11
VATA_PVRAB STANDARD; PRT; 1017 AA.
AC Q9UXU7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE V-type ATP synthase alpha chain (EC 3.6.3.14) (V-type ATPase subunit
DE A) [Contains: Pab atpA intein (Pab VMA intein)].
GN ATPA OR PYRAB17610 OR PAB2378.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GES / Orsay;
RX MEDLINE=22511545; PubMed=12622808;
RA Cohen G.N., Barbe V., Flament D., Galperin M., Heilig R., Lecompte O.,
RA Poch O.G., Prieur D., Querellou J., Ripp R., Thierry J.-C.,
RA Van der Oost J., Weissenbach J., Zivanovic Y., Forterre P.;
RT "An integrated analysis of the genome of the hyperthermophilic
RT archaeon Pyrococcus abyssi.";
RL Mol. Microbiol. 47:1495-1512(2003).
CC -!- FUNCTION: Produces ATP from ADP in the presence of a proton
CC gradient across the membrane. The archaeal alpha chain is a
CC catalytic subunit.
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
CC H(+) (Out).
CC -!- PTM: This protein undergoes a protein self splicing that involves

```

```

CC a post-translational excision of the VDE intervening region
CC (intein) followed by peptide ligation (Potential).
CC -!- MISCELLANEOUS: The intein interrupts the ATP-binding site.
CC -!- SIMILARITY: Belongs to the ATPase alpha/beta chains family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ248288; CAB50666.1; -.
DR PIR; D75028; D75028.
DR HAMAP; MF_00309; atypical; 1.
DR InterPro; IPR000793; ATPase_a/b C.
DR InterPro; IPR000194; ATPase_a/bcentre.
DR InterPro; IPR004100; ATPase_a/bN.
DR InterPro; IPR009005; F1_ATPase_a/bN.
DR InterPro; IPR003587; Hedgehog_hint_N.
DR InterPro; IPR003586; Hedgehog_hintC.
DR InterPro; IPR006142; INTEIN.
DR InterPro; IPR004042; Intein_endonuc.
DR InterPro; IPR006141; Intein_S.
DR Pfam; PF00006; ATP-synt_ab; 1.
DR Pfam; PF00306; ATP-synt_ab_C; 1.
DR Pfam; PF02874; ATP-synt_ab_N; 1.
DR PRINTS; PR00379; INTEIN.
DR SMART; SM00305; HintC; 1.
DR SMART; SM00306; HintN; 1.
DR TIGRFAMs; TIGR01443; intein_Cterm; 1.
DR TIGRFAMs; TIGR01445; intein_Nterm; 1.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
DR PROSITE; PS50818; INTEIN_C_TER; 1.
DR PROSITE; PS50819; INTEIN_N_TER; 1.
DR Hydroxylase; ATP synthase; Hydrogen ion transport; ATP-binding;
KW Autocatalytic cleavage; Protein splicing; Complete proteome.
FT CHAIN 1 240 V-TYPE ATP SYNTHASE ALPHA CHAIN, 1ST
FT CHAIN 241 669 PAB ATPA INTEIN (POTENTIAL).
FT CHAIN 670 1017 V-TYPE ATP SYNTHASE ALPHA CHAIN, 2ND
FT CHAIN PART (POTENTIAL).
SQ SEQUENCE 1017 AA; 114354 MW; 0P200895F3E407A CRC64;

Query Match 43.4%; Score 43; DB 1; Length 1017;
Best Local Similarity 58.8%; Pred. No. 73;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 QRAVRLALSGSDGRVGS 18
Db :|||: |||||
798 ERAGRVTGLSDYRVGS 814

RESULT 12
POLG_BSTV1 STANDARD; PRT; 3093 AA.
AC Q65730;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Genome polyprotein [Contains: N-terminal protein (P1); Helper
DE component proteinase (EC 3.4.22.45) (HC-Pro); Protein p3; 6 kDa
DE protein 1 (6k); Cytoplasmic inclusion protein (CI); 6 kDa protein 2
DE (6k2); Genome-linked protein (VFG); Nuclear inclusion protein A (NI-A)
DE (NIA)-/ (EC 3.4.22.44) (49 kDa proteinase) (49 kDa-Pro); Nuclear
DE inclusion protein B (NI-B) (NIB) (RNA-directed RNA polymerase)
DE (EC 2.7.7.48); Coat protein (CP)].
OS Brome streak virus (strain 11-Cal) (BstV) (Brome streak mosaic
OS tymovirus).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potyviridae;
OC Tritimovirus.

```

OX NCBI_TaxID=117138;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95363423; PubMed=7636484;
RA Goetz R., Maies E.;
RT "The complete nucleotide sequence and genome organization of the
RT mite-transmitted brone streak mosaic rymovirus in comparison with
RT those of potyviruses.";
RL J. Gen. Virol. 76:2035-2035 (1995).
CC -!- FUNCTION: Helper component-proteinase is required for aphid
CC transmission and also has proteolytic activity.
CC -!- FUNCTION: Cytoplasmic inclusion protein has helicase activity. It
CC may be involved in replication.
CC -!- FUNCTION: Nuclear inclusion protein A has proteolytic activity.
CC -!- CATALYTIC ACTIVITY: Hydrolyzes glutamyl bonds, and activity is
CC further restricted by preferences for the amino acids in P6 - P1'
CC that vary with the species of potyvirus, e.g. Glu-Xaa-Xaa-Tyr-Xaa-
CC Glu-(Ser or Gly) for the enzyme from tobacco etch virus. The
CC natural substrate is the viral polyprotein, but other proteins and
CC oligopeptides containing the appropriate consensus sequence are
CC also cleaved.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA}(N).
CC -!- CATALYTIC ACTIVITY: Hydrolyzes a Gly-|-Gly bond at its own C-
CC terminus, commonly in the sequence -Tyr-Xaa-Val-Gly-|-Gly, in the
CC processing of the potyviral polyprotein.
CC -!- PTM: VPG is covalently linked to the genomic RNA.
CC -!- PTM: The viral RNA of potyviruses is expressed as a single
CC polyprotein which undergoes posttranslational proteolytic
CC processing resulting in the production of at least eight
CC individual proteins.
CC -!- SIMILARITY: HC PROTEINASE BELONGS TO PEPTIDASE FAMILY C6.
CC -!- SIMILARITY: NI-A PROTEINASE BELONGS TO PEPTIDASE FAMILY C4.
CC -!- SIMILARITY: Belongs to the potyviruses polyprotein family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z48506; CAA88417.1; -;
CC MEROPS; C04.UFW; -;
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR001730; Peptidase_C4.
DR InterPro; IPR001592; Poty coat.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR001205; RNA_pol_PSD.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00863; Peptidase_C4; 1.
DR Pfam; PF00767; Poty coat; 1.
DR Pfam; PF00680; RNA dep RNA pol; 1.
DR PRINTS; PR00966; NIAPOTYPASE.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 1.
DR Hydrolase; Transferase; Thiol protease; RNA-directed RNA polymerase;
KW Coat protein; Polyprotein; Covalent protein-RNA linkage; Helicase;
KW ATP-binding.
FT CHAIN 1 403 N-TERMINAL PROTEIN.
FT CHAIN 404 809 HELPER COMPONENT PROTEINASE.
FT CHAIN 810 1087 PROTEIN P3.
FT CHAIN 1088 1138 6 kDa PROTEIN 1.
FT CHAIN 1139 1783 CYTOPLASMIC INCLUSION PROTEIN.
FT CHAIN 1784 1834 6 kDa PROTEIN 2.
FT CHAIN 1835 2040 GENOME-LINKED PROTEIN.
FT CHAIN 2041 2275 NUCLEAR INCLUSION PROTEIN A.
FT CHAIN 2276 2827 NUCLEAR INCLUSION PROTEIN B.
FT CHAIN 2828 3093 COAT PROTEIN.
FT BINDING 1915 1915 COVALENT LINKAGE OF VIRAL RNA

FT NP_BIND 1228 1235 (BY SIMILARITY).
FT ATP (POTENTIAL).
SQ SEQUENCE 3093 AA; 348101 MW; 50F775CEFF7ABCCB4 CRC64;

Query Match 43.4%; Score 43; DB 1; Length 3093;
Best Local Similarity 47.1%; Pred. NO. 2.4e+02;
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 4 AVRLALSGSDGRVSGC 20
DB 981 AYRLTYTGVGARIGRC 997

RESULT 13
YHDP ECOLI
ID YHDP_ECOLI STANDARD; PRT; 986 AA.
AC P46474; P46475; P76676;
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein yhdp.
GN YHDP OR B3245.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12.";
RT Science 277:1453-1474 (1997).
RL
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U18997; AAA58047.1; ALT_FRAME.
DR EMBL; U18997; AAA58048.1; ALT_FRAME.
DR EMBL; AB000403; AAC76277.1; -;
DR PIR; G65116; G65116.
DR EcoGene; EG12827; yhdp.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 926 946 POTENTIAL.
SQ SEQUENCE 986 AA; 107687 MW; 3D0EFF9696724ACD CRC64;

Query Match 42.9%; Score 42.5; DB 1; Length 986;
Best Local Similarity 60.0%; Pred. No. 84;
Matches 12; Conservative 1; Mismatches 4; Indels 3; Gaps 1;

QY 1 LQRAVRLALSGS---DGRVG 17
DB 477 LPEAVNEALSGSVAWDGVKVG 496

RESULT 14
ROC_HUMAN
ID ROC_HUMAN STANDARD; PRT; 306 AA.
AC P07910; P22628; Q96HK7; Q96HM4; Q96IY5; Q9BTS3;
DT 01-AUG-1988 (Rel. 08, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Heterogeneous nuclear ribonucleoproteins C1/C2 (hnRNP C1 / hnRNP C2).
GN HNRPC.
OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;
 [1]
 RP SEQUENCE FROM N.A. (ISOFORM C2).
 RX MEDLINE=90099350; PubMed=2557628;
 RA Burd C.G., Swanson M.S., Goerlach M., Dreyfuss G.;
 RT "Primary structures of the heterogeneous nuclear ribonucleoprotein
 RT A2, B1, and C2 proteins: a diversity of RNA binding proteins is
 RT generated by small peptide inserts.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:9788-9792(1989).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM C1).
 RX MEDLINE=87257872; PubMed=3110598;
 RA Swanson M.S., Nakagawa T.Y., Levay K., Dreyfuss G.;
 RT "Primary structure of human nuclear ribonucleoprotein particle C
 RT proteins: conservation of sequence and domain structures in
 RT heterogeneous nuclear RNA, mRNA, and pre-rRNA-binding proteins.";
 RL Mol. Cell. Biol. 7:1731-1739(1987).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORMS C1 AND C2).
 RC TISSUE=Bone marrow, Brain, Eye, and Placenta;
 RX MEDLINE=23388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Uudin T.B., Tothiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S.C., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [4]
 RP PARTIAL SEQUENCE, AND CHARACTERIZATION.
 RX MEDLINE=90067819; PubMed=2587210;
 RA Merrill B.M., Barnett S.F., Lestourgeon W.M., Williams K.R.;
 RT "Primary structure differences between proteins C1 and C2 of HeLa 40S
 RT nuclear ribonucleoprotein particles.";
 RL Nucleic Acids Res. 17:8441-8449(1989).
 RN [5]
 RP STRUCTURE BY NMR OF 1-94.
 RX MEDLINE=92329450; PubMed=1385725;
 RA Witteking M., Goerlach M., Friedrichs M., Dreyfuss G., Mueller L.;
 RT "1H, 13C, and 15N NMR assignments and global folding pattern of the
 RT RNA-binding domain of the human hnRNP C proteins.";
 RL Biochemistry 31:6254-6265(1992).
 RN [6]
 RP STRUCTURE BY NMR OF 1-94.
 RX MEDLINE=92371436; PubMed=1380452;
 RA Goerlach M., Witteking M., Beckman R.A., Mueller L., Dreyfuss G.;
 RT "Interaction of the RNA-binding domain of the hnRNP C proteins with
 RT RNA.";
 RL EMBO J. 11:3289-3295(1992).
 CC -!- FUNCTION: May play a role in ribonucleosome assembly by
 CC neutralizing basic proteins such as A and B core hnRNP.
 CC -!- SUBCELLULAR LOCATION: Nuclear; component of ribonucleosomes.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=C2;
 CC IsoId=P07910-1; Sequence=Displayed;
 CC Name=C1;
 CC IsoId=P07910-2; Sequence=VSP_005831;

CC -!- PTM: Phosphorylated (Probable).
 CC -!- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M29063; AAA36576.1; -;
 DR EMBL: M16342; AAA52680.1; -;
 DR EMBL: BC003394; AAH03394.1; -;
 DR EMBL: BC007052; AAH07052.1; -;
 DR EMBL: BC008364; AAH08364.1; -;
 DR EMBL: BC008423; AAH08423.1; -;
 DR PIR: A26885; A26885.
 DR PIR: C34504; C34504.
 DR SWISS-2DPAGE; P07910; HUMAN.
 DR Aarhus/Ghent-2DPAGE; 7207; IEF.
 DR Aarhus/Ghent-2DPAGE; 7222; IEF.
 DR Genew: HGNC:5035; HNRPC.
 DR GK; P07910; -;
 DR MIM; 164020; -;
 DR GO; GO:0005634; C:nucleus; TAS.
 DR GO; GO:0006371; P:mRNA splicing; TAS.
 DR InterPro: IPR000504; RNA_rec_mot.
 DR Pfam; PF00076; xtm; 1.
 DR SMART; SM00360; RRM; 1.
 DR PROSITE; PS0102; RRM; 1.
 DR PROSITE; PS00303; RRM RNP 1; 1.
 KW Nuclear protein; RNA-binding; Ribonucleoprotein; Phosphorylation;
 KW Alternative splicing.
 FT DOMAIN 16 87 RNA-BINDING (RRM).
 FT DOMAIN 155 161 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 FT DOMAIN 181 303 ASP/GLU-RICH (ACIDIC).
 FT VARSPLIC 108 120 Missing (in isoform C1).
 FT CONFLICT 110 110 E -> G (IN REF. 3; AAH08423).
 FT CONFLICT 217 217 M -> I (IN REF. 3).
 FT CONFLICT 224 224 K -> R (IN REF. 3; AAH08364).
 FT CONFLICT 244 244 K -> R (IN REF. 3; AAH08364).
 FT CONFLICT 254 254 E -> G (IN REF. 3; AAH07052).
 FT CONFLICT 303 306 EDDS -> G (IN REF. 1 AND 2).
 SQ SEQUENCE 306 AA; 33688 MW; 17BF8382D0829818 CRC64;
 Query Match 42.4%; Score 42; DB 1; Length 306;
 Best Local Similarity 44.4%; Pred. No. 28;
 Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;
 QY 2 QRAVRLALSGSDGRVCSG 19
 Db 60 ERNARAAGEDGRMIAG 77
 RESULT 15
 ROC_MOUSE
 ID ROC_MOUSE STANDARD; PRT; 313 AA.
 AC Q92204; Q9CQT3; Q9CY83;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Heterogeneous nuclear ribonucleoproteins C1/C2 (hnRNP C1 / hnRNP C2).
 GN HNRPC OR HNRNPC.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM C2).
 RX MEDLINE=20266297; PubMed=10805751;
 RA Williamson D.J., Banik-Maiti S., DeGregori J., Ruley H.E.;


```

KW rRNA processing; Ribosome biogenesis.
FT NP_BIND 51 58 ATP (POTENTIAL).
FT SITE 166 169 DEAD BOX.
SQ SEQUENCE 546 AA; 61211 MW; 3C19F7354C29F9A1 CRC64;

Query Match 42.4%; Score 42; DB 1; Length 546;
Best Local Similarity 45.5%; Pred. No. 53;
Matches 10; Conservative 3; Mismatches 5; Indels 4; Gaps 1;

QY 2 QRAVRLALSGSD----GRVSGS 19
Db 35 EKAIPLALEGKOLLARARTGSG 56

RESULT 17
ID DX56 HUMAN STANDARD; PRT; 547 AA.
AC Q9NY93; Q9H918;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable ATP-dependent 61 kDa nucleolar RNA helicase (BC 3.6.1.-)
DE (DEAD-box protein 56) (DEAD-box protein 21).
GN DDX56 OR NOH61 OR DDX21.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_TaxID=9606;
RP SEQUENCE FROM N.A.
RC TISSUE=Keratinocytes; PubMed=10749921;
RX MEDLINE=20214812; PubMed=10749921;
RA Zirmes R.F., Eilbracht J., Kneissel S., Schmidt-Zachmann M.S.;
RT "A novel helicase-type protein in the nucleolus: protein NOH61."
RL Mol. Biol. Cell 11:1153-1167(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Camargo A.A., Nunes D.N., Samaha H.B., Simpson A.J.G., Dias-Neto E.;
RT "Sequencing of a new human DEAD-box RNA helicase."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=21154917; PubMed=11230166;
RA Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,
RA Ansoerge W., Boecher M., Bloecker H., Bauersachs S., Blum H.,
RA Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,
RA Mewes H.-W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,
RA Wambutt R., Korn B., Klein M., Pousta A.;
RT "Towards a catalog of human genes and proteins: sequencing and
RT analysis of 500 novel complete protein coding human cDNAs."
RL Genome Res. 11:422-435(2001).
RN [4]
RP SEQUENCE FROM N.A.
RA Isozaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Takatsuna M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakanatsu A., Nakamura Y., Nagahara K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NED0 human cDNA sequencing project."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC TISSUE=Cervix;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

```

```

RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [6]
RP SUBCELLULAR LOCATION.
RX MEDLINE=22317277; PubMed=12429849;
RA Scherl A., Coute Y., Deon C., Calle A., Kindbeiter K., Sanchez J.-C.,
RA Greco A., Hochstrasser D.F., Diaz J.-J.;
RT "Functional proteomic analysis of human nucleolus."
RL Mol. Biol. Cell 13:4100-4109(2002).
CC -!- FUNCTION: May play a role in later stages of the processing of the
CC pre-ribosomal particles leading to mature 60S ribosomal subunits.
CC Has intrinsic ATPase activity.
CC -!- SUBUNIT: May form homo-oligomeric complexes.
CC -!- SUBCELLULAR LOCATION: Nuclear; nucleolar.
CC -!- TISSUE SPECIFICITY: Detected in heart, brain, liver, pancreas,
CC placenta and lung.
CC -!- SIMILARITY: Belongs to the DEAD box helicase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ131712; CAB87992.1; -
DR EMBL; AF247666; AAG36876.1; -
DR EMBL; AL136700; CAB66635.1; -
DR EMBL; AK022774; BAB14238.1; -
DR EMBL; BC001235; AAH01235.1; -
DR SWISS-2DPAGE; Q9NY93; HUMAN.
DR Genew; HGNC:18193; DDX56.
DR GK; Q9NY93; -
DR MIM; 608023; -
DR GO; GO:0005730; C:nucleolus; TAS.
DR GO; GO:0004004; F:ATP dependent RNA helicase activity; TAS.
DR GO; GO:0007046; P:ribosome biogenesis; TAS.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00271; Helicase_C; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 1.
KW Hydrolase; Helicase; ATP-binding; Nuclear protein;
KW rRNA processing; Ribosome biogenesis.
FT NP_BIND 51 58 ATP (POTENTIAL).
FT SITE 166 169 DEAD BOX.
FT CONFLICT 145 145 I -> T (IN REF. 4).
SQ SEQUENCE 547 AA; 61589 MW; 6D127CDA73610D2 CRC64;

Query Match 42.4%; Score 42; DB 1; Length 547;
Best Local Similarity 45.5%; Pred. No. 53;
Matches 10; Conservative 3; Mismatches 5; Indels 4; Gaps 1;

QY 2 QRAVRLALSGSD----GRVSGS 19
Db 35 EKAIPLALEGKOLLARARTGSG 56

RESULT 18

```

```

VATA PYRAE
ID VATA PYRAE STANDARD; PRT; 593 AA.
AC Q8ZVL1.
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE V-type ATP synthase alpha chain (EC 3.6.3.14) (V-type ATPase subunit
DE A).
GN ATPA OR PAE0663.
OS Pyrobaculum aerophilum.
OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
OC Thermoproteaceae; Pyrobaculum.
OX NCBI_TaxID=13773;
RN [1] SEQUENCE FROM N.A.
RC STRAIN=FM2 / ATCC 51768 / DSM 7523;
RX MEDLINE=21664397; PubMed=11792869;
RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
RA Miller J.H.;
RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
RT aerophilum."
RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
CC -!- FUNCTION: Produces ATP from ADP in the presence of a proton
CC gradient across the membrane. The archaeal alpha chain is a
CC catalytic subunit.
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
CC H(+) (Out).
CC -!- SIMILARITY: Belongs to the ATPase alpha/beta chains family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AF009777; AL62932.1; -.
DR HAMAP; MF_00309; -.
DR InterPro; IPR000793; ATPase_a/b_C.
DR InterPro; IPR000194; ATPase_a/b_Centre.
DR InterPro; IPR004100; ATPase_a/bN.
DR InterPro; IPR009005; F1_ATPase_a/bN.
DR Pfam; PF00006; ATP-synt_ab; 1.
DR Pfam; PF00306; ATP-synt_ab_C; 1.
DR Pfam; PF02874; ATP-synt_ab_N; 1.
DR PROSITE; PS00152; ATPASE ALPHA_BETA; FALSE NEG.
DR Hydrolase; ATP synthetase; Hydrogen ion transport; ATP-binding;
KW Complete proteome.
FT NP_BIND 236 243 ATP (POTENTIAL).
FT SEQUENCE 593 AA; 66319 MW; C3966BF19D7FBC3 CRC64;
SQ
Query Match 42.4%; Score 42; DB 1; Length 593;
Best Local Similarity 58.8%; Pred. No. 58;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
QY 2 QRAVRLALSGDGRVGS 18
DB 371 ERAGRVLYGSKERVGS 387
RESULT 19
YORI_ YEAST
ID YORI_ YEAST STANDARD; PRT; 1477 AA.
AC P53049;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Oligomycin resistance ATP-dependent permease YORI1.
DE YORI1 OR YGR281W.
GN Saccharomyces cerevisiae (Baker's yeast).
OS Saccharomycetes; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Eukaryota; Fungi; Ascomycota; Saccharomycetales; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1] SEQUENCE FROM N.A.
RC STRAIN=S288c / FY1679;
RX MEDLINE=96069397; PubMed=8524254;
RA Katzmann D.J., Hallstrom T.C., Voet M., Wysock W., Golin J.,
RA Volkmaert G., Moye-Rowley W.S.;
RT "Expression of an ATP-binding cassette transporter-encoding gene
RT (YOR1) is required for oligomycin resistance in Saccharomyces
RT cerevisiae."
RL Mol. Cell. Biol. 15:6875-6883(1995).
RN [2] SEQUENCE FROM N.A.
RC STRAIN=S288c / FY1679;
RX MEDLINE=97245295; PubMed=9090054;
RA Volkmaert G., Voet M., Robben J.;
RT "Sequence analysis of a near-subtelomeric 35.4 kb DNA segment on the
RT right arm of chromosome VII from Saccharomyces cerevisiae carrying
RT the MAL1 locus reveals 15 complete open reading frames, including
RT ZUO1, BGL2 and BIO2 genes and an ABC transporter gene."
RL Yeast 13:251-259(1997).
CC -!- FUNCTION: Required for oligomycin resistance.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: Belongs to the ABC transporter family. MDR subfamily.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
EMBL; Z73066; CA997312.1; -.
DR FIR; S84616; S64616.
DR HSSP; P13569; INBD.
DR GerMOnline; 141593; -.
DR SGD; S0003513; YORI1.
DR GO; GO:0005886; C:plasma membrane; IDA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR001140; ABC_TW_Transpt.
DR InterPro; IPR003439; ABC_Transporter.
DR Pfam; PF00664; ABC_membrane; 2.
DR Pfam; PF00005; ABC_tran; 2.
DR ProDom; PD000006; ABC_transporter; 2.
DR SMART; SM00382; AAA_2.
DR PROSITE; PS00929; ABC_TWIF; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 2.
KW ATP-binding; Transmembrane; Glycoprotein; Transport.
FT TRANSMEM 207 227 POTENTIAL.
FT TRANSMEM 250 270 POTENTIAL.
FT TRANSMEM 329 349 POTENTIAL.
FT TRANSMEM 350 370 POTENTIAL.
FT TRANSMEM 434 454 POTENTIAL.
FT TRANSMEM 479 499 POTENTIAL.
FT TRANSMEM 616 636 POTENTIAL.
FT TRANSMEM 893 913 POTENTIAL.
FT TRANSMEM 941 961 POTENTIAL.
FT TRANSMEM 1028 1048 POTENTIAL.
FT TRANSMEM 1118 1138 POTENTIAL.
FT TRANSMEM 1142 1162 POTENTIAL.
FT NP_BIND 621 628 ATP (POTENTIAL).
FT NP_BIND 1247 1254 ATP (POTENTIAL).
FT CARBOHYD 16 16 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 295 295 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 661 661 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 759 759 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 799 799 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1345 1345 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1366 1366 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1477 AA; 166727 MW; 40C5D36CA9B6A8C5 CRC64;
Query Match 42.4%; Score 42; DB 1; Length 1477;

```



```
Best Local Similarity 42.1%; Pred. No. 1.6e+02;
Matches 8; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

QY 1 LORAVRLALSGSGRGVSG 19
DB 772 IERASRVVLGTQGVQVDIG 790

RESULT 20
TOR2 SCHPO
ID TOR2 SCHPO STANDARD; PRT; 2337 AA.
AC Q9Y7Z2; O94507;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Phosphatidylinositol 3-kinase tor2 (EC 2.7.1.137) (PI3-kinase)
DE (PtdIns-3-kinase) (PI3K)
GN TOR2 OR SPBC216.07C OR SPBC646.01C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neil S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vansteels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.B., Moreno S., Armstrong J., Forburg S.L.,
RA Cerretti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
RN [2]
RN IDENTIFICATION.
RX MEDLINE=21269225; PubMed=11096119;
RA Weisman R., Choder M.;
RT "The fission yeast TOR homolog, tor1+, is required for the response to
RT starvation and other stresses via a conserved serine.";
RL J. Biol. Chem. 276:7027-7032(2001).
CC -1- FUNCTION: Phosphatidylinositol 3-kinase homolog required for G1
CC progression (By similarity).
CC -1- CATALYTIC ACTIVITY: ATP + 1-phosphatidyl-1D-myo-inositol = ADP +
CC 1-phosphatidyl-1D-myo-inositol 3-phosphate.
CC -1- SIMILARITY: Belongs to the PI3/P14-kinase family.
CC -1- SIMILARITY: Contains 15 HEAT repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
CC

or send an email to license@isb-sib.ch).
EMBL; AL035216; CAA22805.1; -
DR EMBL; AL049558; CAB40167.1; -
DR PIR; T39913; T40577.
DR HSSP; P42345; LAUE.
DR GenedB_Spombe; SPEC216.07c; -.
DR InterPro; IPR008938; ARM.
DR InterPro; IPR003151; FAT.
DR InterPro; IPR003152; FATC.
DR InterPro; IPR000357; HEAT.
DR InterPro; IPR000403; PI3_P14_kinase.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF02259; FAT; 1.
DR Pfam; PF00454; PI3_P14_kinase; 1.
DR SMART; SM00146; PI3K; 1.
DR PROSITE; PS00777; HEAT REPEAT; 1.
DR PROSITE; PS00915; PI3_4_KINASE_1; 1.
DR PROSITE; PS00916; PI3_4_KINASE_2; FALSE_NEG.
DR PROSITE; PS0290; PI3_4_KINASE_3; 1.
KW Transferase; Kinase; Cell cycle; Repeat.
FT REPEAT 163 200 HEAT 1.
FT REPEAT 249 286 HEAT 2.
FT REPEAT 290 326 HEAT 3.
FT REPEAT 409 446 HEAT 4.
FT REPEAT 474 512 HEAT 5.
FT REPEAT 559 596 HEAT 6.
FT REPEAT 642 679 HEAT 7.
FT REPEAT 683 721 HEAT 8.
FT REPEAT 727 765 HEAT 9.
FT REPEAT 802 840 HEAT 10.
FT REPEAT 844 880 HEAT 11.
FT REPEAT 881 921 HEAT 12.
FT REPEAT 965 1004 HEAT 13.
FT REPEAT 1006 1043 HEAT 14.
FT REPEAT 1735 1772 HEAT 15.
FT DOMAIN 1990 2337 PI3K/PI4K.
SQ SEQUENCE 2337 AA; 266376 MW; 190F448DA04FD2D9 CRC64;

Query Match 42.4%; Score 42; DB 1; Length 2337;
Best Local Similarity 57.1%; Pred. No. 2.6e+02;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 QRAVRLALSGSGR 15
DB 1969 QRPRRLSIKSGDOK 1982

RESULT 21
RECG HAEIN
ID RECG HAEIN STANDARD; PRT; 693 AA.
AC P43809;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE ATP-dependent DNA helicase recG (EC 3.6.1.-).
GN RECG OR H11740
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
```

```

RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
CC -!- FUNCTION: Critical role in recombination and DNA repair. Help
CC process Holliday junction intermediates to mature products by
CC catalyzing branch migration. Has a DNA unwinding activity
CC characteristic of a DNA helicase with a 3' to 5' polarity. RecG
CC unwind branched duplex DNA (Y-DNA) (By similarity).
CC -!- SIMILARITY: Belongs to the helicase family. RecG subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U32847; AAC23387.1; -.
DR FIRM; E64139; E64139.
DR TIGR; H11740; -.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR008994; Nucleic_acid_OB.
DR InterPro; IPR004609; RecG.
DR InterPro; IPR004365; CRNA_anti.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF01336; trna_anti; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
DR TIGRfam; TIGR00643; recG; 1.
KW Hydrolase; Helicase; DNA repair; ATP-binding; DNA recombination;
KW DNA-binding; Complete proteome.
FT NP_BIND 296 303 ATP (POTENTIAL).
FT SITE 397 400 DEQH BOX.
FT SEQUENCE 693 AA; 78085 MW; 12D1BC65F0259950 CRC64;
Query Match 41.9%; Score 41.5; DB 1; Length 693;
Best Local Similarity 47.8%; Pred. No. 83;
Matches 11; Conservative 2; Mismatches 7; Indels 3; Gaps 1;
QY 1 LQRAVRLALSGSD---GRVSGC 20
DB 567 IENAERLGLSQLHQRVGRGC 589
RESULT 22
YHHK_ECOLI
ID YHHK_ECOLI STANDARD; PRT; 127 AA.
AC P37613;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein yhhk.
GN YHHK OR B3459.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=94316500; PubMed=8041620;
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT region from 76.0 to 81.5 minutes."
RL Nucleic Acids Res. 22:2576-2586(1994).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U00039; AAB18434.1; -.
DR EMBL; AE000422; AAC76484.1; -.
DR PIR; S47678; S47678.
DR EcoGene; EG12211; yhhk.
DR InterPro; IPR000182; GCN5acetyl_trans.
DR Pfam; PF00583; Acetyltransf; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 127 AA; 14505 MW; EF054A30D294519A CRC64;
Query Match 41.4%; Score 41; DB 1; Length 127;
Best Local Similarity 50.0%; Pred. No. 16;
Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
QY 1 LQRAVRLALSGSDGRVGS 18
DB 48 LLAARVTLSTEGALDS 65
RESULT 23
CDD_MOUSE
ID CDD_MOUSE STANDARD; PRT; 146 AA.
AC P56389;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Cytidine deaminase (EC 3.5.4.5) (Cytidine aminohydrolase).
GN CDA OR CDD.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Marra M., Hillier L., Allen M., Bowles M., Dietrich N., Dubuque T.,
RA Geisel S., Kucaba T., Lacy M., Le M., Martin J., Morris M.,
RA Schellenberg K., Steptoe M., Tan F., Underwood K., Moore B.,
RA Theising B., Wylie T., Lennon G., Soares B., Wilson R., Waterston R.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: This enzyme scavenges exogenous and endogenous cytidine
CC and 2'-deoxycytidine for UMP synthesis (By similarity).
CC -!- CATALYTIC ACTIVITY: Cytidine + H(2)O = uridine + NH(3).
CC -!- COFACTOR: Zinc (By similarity).
CC -!- SUBUNIT: Homotrimer (By similarity).
CC -!- SIMILARITY: Belongs to the cytidine and deoxycytidylate deaminases
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AA388666; -. NOT_ANNOTATED_CDS.
DR HSSP; P13652; 1CTT.
DR InterPro; IPR006262; Cyt_deam_tetra.
DR InterPro; IPR002125; dCMP_cyt_deam.
DR Pfam; PF00383; dCMP_cyt_deam; 1.
DR TIGRfam; TIGR01354; cyt_deam_tetra; 1.
DR PROSITE; PS00903; CYT_DCMP_DEAMINASES; 1.
KW Hydrolase; Zinc.
FT METAL 65 65 ZINC (BY SIMILARITY).
FT METAL 99 99 ZINC (BY SIMILARITY).
FT METAL 102 102 ZINC (BY SIMILARITY).
SQ SEQUENCE 146 AA; 16230 MW; 81B87AB6BABA300A CRC64;

```

```
Query Match 41.4%; Score 41; DB 1; Length 146;
Best Local Similarity 56.2%; Pred. No. 18;
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 5 VRLALSGSDGRVSGC 20
DB 38 VQAALLTGDRIFSGC 53

RESULT 24
DAPB VIBCH STANDARD; PRT; 269 AA.
AC Q9KEH7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dihydrodipicolinate reductase (EC 1.3.1.26) (DHPR).
GN DAPB OR VC2391.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]_TaxID=666;
RP SEQUENCE FROM N.A.
RC STRAIN=EI Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heideberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Seilers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
CC -!- CATALYTIC ACTIVITY: 2,3,4,5-tetrahydrodipicolinate + NAD(P) (+) =
CC 2,3-dihydrodipicolinate + NAD(P)H.
CC -!- PATHWAY: Biosynthesis of diaminoimelate and lysine from aspartate
CC semialdehyde, second step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the dihydrodipicolinate reductase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF004310; AAF95534.1; -.
CC FIC; C82080; C82080.
CC HSP; P04036; IDRW.
CC TIGR; VC2391; -.
CC HAMAP; MF_00102; -.
CC InterPro; IPR000846; DapB.
CC Pfam; PF05173; DapB_C; 1.
CC Pfam; PF01113; DapB_N; 1.
CC ProDom; PD004105; DapB; 1.
CC TIGRFAMs; TIGR00036; dapB; 1.
CC PROSITE; PS01298; DAPB; 1.
CC Diaminopimelate biosynthesis; Lysine biosynthesis; Oxidoreductase;
KW NADP; Complete proteome.
SQ SEQUENCE 269 AA; 5D71DA0CB346EB82 CRC64;

Query Match 41.4%; Score 41; DB 1; Length 269;
Best Local Similarity 53.8%; Pred. No. 36;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 5 VRLALSGSDGRVG 17
DB 2 VRIATAGAGRMG 14

RESULT 26
DCOR_CHICK STANDARD; PRT; 450 AA.
AC DCOR_CHICK;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ornithine decarboxylase (EC 4.1.1.17) (ODC) (Fragment).
GN ODC.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;

Query Match 41.4%; Score 41; DB 1; Length 146;
Best Local Similarity 56.2%; Pred. No. 18;
Matches 9; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 5 VRLALSGSDGRVSGC 20
DB 38 VQAALLTGDRIFSGC 53

RESULT 24
DAPB VIBCH STANDARD; PRT; 269 AA.
AC Q9KEH7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dihydrodipicolinate reductase (EC 1.3.1.26) (DHPR).
GN DAPB OR VC2391.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]_TaxID=666;
RP SEQUENCE FROM N.A.
RC STRAIN=EI Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heideberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Seilers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
CC -!- CATALYTIC ACTIVITY: 2,3,4,5-tetrahydrodipicolinate + NAD(P) (+) =
CC 2,3-dihydrodipicolinate + NAD(P)H.
CC -!- PATHWAY: Biosynthesis of diaminoimelate and lysine from aspartate
CC semialdehyde, second step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the dihydrodipicolinate reductase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF004310; AAF95534.1; -.
CC FIC; C82080; C82080.
CC HSP; P04036; IDRW.
CC TIGR; VC2391; -.
CC HAMAP; MF_00102; -.
CC InterPro; IPR000846; DapB.
CC Pfam; PF05173; DapB_C; 1.
CC Pfam; PF01113; DapB_N; 1.
CC ProDom; PD004105; DapB; 1.
CC TIGRFAMs; TIGR00036; dapB; 1.
CC PROSITE; PS01298; DAPB; 1.
CC Diaminopimelate biosynthesis; Lysine biosynthesis; Oxidoreductase;
KW NADP; Complete proteome.
SQ SEQUENCE 269 AA; 5D71DA0CB346EB82 CRC64;

Query Match 41.4%; Score 41; DB 1; Length 269;
Best Local Similarity 53.8%; Pred. No. 36;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 5 VRLALSGSDGRVG 17
DB 2 VRIATAGAGRMG 14
```

```

RA  SEQUENCE FROM N.A.
RT  "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL  Nature 409:529-533(2001).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=O157:H7 / RIMD 0509952;
RX  MEDLINE=21156231; PubMed=11258796;
RA  Hayashi T., Makino K., Ohnishi M., Kurokawa K., Iehli K., Yokoyama K.,
RA  Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA  Lida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
RA  Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT  "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT  O157:H7 and genomic comparison with a laboratory strain K-12.";
RL  DNA Res. 8:11-22(2001).
CC  -!- FUNCTION: Catalyzes the dehydration of 2-methylcitrate to 2-
CC  methyl-cis-aconitate (By similarity).
CC  -!- CATALYTIC ACTIVITY: 2-hydroxybutane-1,2,3-tricarboxylate = (Z)-
CC  but-2-ene-1,2,3-tricarboxylate + H(2)O.
CC  -!- COFACTOR: Binds 1 2Fe-2S cluster (By similarity).
CC  -!- PATHWAY: Propionate catabolism; 2-methylcitric acid cycle; third
CC  step.
CC  -!- SUBUNIT: Monomer (By similarity).
CC  -!- SIMILARITY: BELONGS TO THE MMGE / PRPD FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AE005212; AAG54682.1; -.
DR  EMBL; AP002551; BAB33810.1; -.
DR  PIR; C90677; C90677.
DR  PIR; F85527; F85527.
DR  InterPro; IPR005656; MMGE Prpd.
DR  Pfam; PF03972; MMGE Prpd_1.
KW  Lyase; Metal-binding; Iron-sulfur; Iron; 2Fe-2S; Complete proteome.
FT  INIT_MET 0
FT  MET 0 BY SIMILARITY.
SQ  SEQUENCE 482 AA; 53820 MW; 19714CD931C08227 CRC64;
Query Match 41.4%; Score 41; DB 1; Length 482;
Best Local Similarity 60.0%; Pred. No. 67;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 3 RAVRLALSGSDGRVG 17
Db 230 RAVRLALMAKTGEMG 244

RESULT 28
PRPD_EC057
ID PRPD_EC057 STANDARD; PRT; 482 AA.
AC Q8X693;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE 2-methylcitrate dehydratase (EC 4.2.1.79).
GN PRPD OR B0334.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";

```

```

RL Science 277:1453-1474 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Duncan M., Allen E., Araujo R., Aparicio A.M., Chung E., Davis K.,
RA Federspiel N., Hyman R., Kallman S., Komp C., Kurdi O., Lew H., Lin D.,
RA Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP CHARACTERIZATION, AND SEQUENCE OF 1-12.
RX MEDLINE=21642584; PubMed=11782506;
RA Blank L., Green J., Guest J.R.;
RT "Acnc of Escherichia coli is a 2-methylcitrate dehydratase (PrpD) that
RT can use citrate and isocitrate as substrates.";
RL Microbiology 148:133-146 (2002).
CC -!- FUNCTION: Catalyzes the dehydration of 2-methylcitrate to 2-
CC methyl-cis-aconitate. Also seems to be responsible for the
CC residual aconitate activity of the acnAB-null strain.
CC -!- CATALYTIC ACTIVITY: 2-hydroxybutane-1,2,3-tricarboxylate = (Z)-
CC but-2-ene-1,2,3-tricarboxylate + H(2)O.
CC -!- COFACTOR: Binds 1 2Fe-2S cluster.
CC -!- PATHWAY: Propionate catabolism; 2-methylcitric acid cycle; third
CC step.
CC -!- SUBUNIT: Monomer.
CC -!- SIMILARITY: BELONGS TO THE MMGE / PRPD FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; A5000140; AAC73437.1; -.
DR EMBL; U73857; AAB18058.1; -.
DR PIR; P64760; P64760.
DR Ecogene; EGI3603; prpD.
DR InterPro; IPR005656; Mmge PrpD.
DR Pfam; PF03972; Mmge PrpD; 1.
KW Lyase; Metal-binding; Iron-sulfur; Iron; 2Fe-2S; Complete proteome.
FT INIT MET 0
FT SEQUENCE 482 AA; 53820 MW; A7AA5CD5391B82C5 CRC64;
SQ
Query Match 41.4%; Score 41; DB 1; Length 482;
Best Local Similarity 60.0%; Pred. No. 67;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 3 RAVRLALSGDGRVG 17
DB 230 RAVRLALMAKTGMG 244
|||||
|:|

RESULT 29
PRPD_SALTY STANDARD; PRT; 483 AA.
AC Q8Z903;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE 2-methylcitrate dehydratase (EC 4.2.1.79).
GN PRPD OR STY0402 OR T2494.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CT18;
RL MEDLINE=21534947; PubMed=11677608;
RA Parhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Hasham D., Brooks K., Chillingworth T., Connerton P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,

```

```

RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Ty2 / ATCC 700931;
RL MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18.";
RL J. Bacteriol. 185:2330-2337 (2003).
CC -!- FUNCTION: Catalyzes the dehydration of 2-methylcitrate to 2-
CC methyl-cis-aconitate (By similarity).
CC -!- CATALYTIC ACTIVITY: 2-hydroxybutane-1,2,3-tricarboxylate = (Z)-
CC but-2-ene-1,2,3-tricarboxylate + H(2)O.
CC -!- COFACTOR: Binds 1 2Fe-2S cluster (By similarity).
CC -!- PATHWAY: Propionate catabolism; 2-methylcitric acid cycle; third
CC step.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SIMILARITY: BELONGS TO THE MMGE / PRPD FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL627266; CAD08825.1; -.
DR EMBL; AE016842; AAO70082.1; -.
DR InterPro; IPR005656; Mmge PrpD.
DR Pfam; PF03972; Mmge PrpD; 1.
KW Lyase; Metal-binding; Iron-sulfur; Iron; 2Fe-2S; Complete proteome.
FT CONFLICT 452 452 L -> P (IN REF. 2).
FT SEQUENCE 483 AA; 53769 MW; 4C695F381E7E228A CRC64;
SQ
Query Match 41.4%; Score 41; DB 1; Length 483;
Best Local Similarity 60.0%; Pred. No. 67;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 3 RAVRLALSGDGRVG 17
DB 231 RAVRLALMAKTGMG 245
|||||
|:|

RESULT 30
PRPD_SALTY STANDARD; PRT; 483 AA.
AC P74840;
DT 01-NOV-1997 (Rel. 35, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE 2-methylcitrate dehydratase (EC 4.2.1.79).
GN PRPD OR STW0370.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=LT2;
RL MEDLINE=97158691; PubMed=9006051;
RA Horowitz A.R., Escalante-Semerena J.C.;
RT "Propionate catabolism in Salmonella typhimurium LT2: two divergently
RT transcribed units comprise the prp locus at 8.5 centisomes, prpR
RT encodes a member of the sigma-54 family of activators, and the
RT prpBCDE genes constitute an operon.";

```

```

RL J. Bacteriol. 179:928-940(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856(2001).
RN [3]
RP FUNCTION, AND PATHWAY.
RX MEDLINE=21191730; PubMed=11294638;
RA Horswill A.R., Escalante-Semerena J.C.;
RT "In vitro conversion of propionate to pyruvate by Salmonella enterica
RT enzymes: 2-methylcitrate dehydratase (prpD) and aconitase enzymes
RT catalyze the conversion of 2-methylcitrate to 2-methylisocitrate.";
RL Biochemistry 40:4703-4713(2001).
CC -!- FUNCTION: Catalyzes the dehydration of 2-methylcitrate to 2-
CC methyl-cis-aconitate.
CC -!- CATALYTIC ACTIVITY: 2-hydroxybutane-1,2,3-tricarboxylate = (Z)-
CC but-2-ene-1,2,3-tricarboxylate + H(2)O.
CC -!- COFACTOR: Binds 1 2Fe-2S Cluster (By similarity).
CC -!- PATHWAY: Propionate catabolism; 2-methylcitric acid cycle; third
CC step.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SIMILARITY: BELONGS TO THE MMGE / PRPD FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U51879; AAC44816.1; -.
DR EMBL; AF008712; AAL19324.1; -.
DR StyGene; SG10688; prpD.
DR InterPro; IPR005656; Mmge PrpD.
DR Pfam; PF03972; Mmge_Prpd_1.
KW Lyase; Metal-binding; Iron-sulfur; Iron; 2Fe-2S; Complete proteome.
FT CONFLICT 308 308 Q -> H (IN REF. 1).
SQ SEQUENCE 483 AA; 53786 MW; CF045F6958157467 CRC64;
Query Match 41.4%; Score 41; DB 1; Length 483;
Best Local Similarity 60.0%; Pred. No. 67;
Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
QY 3 RAVRLALSGSDGRVG 17
Db |||||:|
231 RAVRLALMAKTGMG 245

```

Search completed: May 13, 2004, 06:55:49
Job time : 8.66667 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 13, 2004, 06:49:47 ; Search time 33.3333 Seconds
(without alignments)
189.311 Million cell updates/sec

Title: 09549186-9

Perfect score: 99
Sequence: 1 LQRAVRLALSGDGRVSGC 20

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : SPREMBL_25.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	52	52.5	373	10 Q7XKP6	Q7xkp6 oryza sativ
2	52	52.5	379	10 Q7XMO0	Q7xm00 oryza sativ
3	52	52.5	405	10 Q9AYM0	Q9aym0 oryza sativ
4	52	52.5	405	10 Q7X976	Q7x976 oryza sativ
5	49.5	50.0	1482	16 Q8EHA4	Q8eha4 shewanella
6	49	49.5	575	17 Q28180	Q28180 archaeoglob
7	48	48.5	1865	11 Q60544	Q60544 mesocricetu
8	47	47.5	114	1 Q8U4U3	Q8u4u3 methanosarc
9	47	47.5	117	17 Q8PYW7	Q8pyw7 methanosarc
10	46.5	47.0	325	2 Q8RMH1	Q8rmh1 acetobacter
11	46	46.5	352	16 Q92AC6	Q92ac6 listeria in
12	46	46.5	352	16 Q8Y619	Q8y619 listeria mo
13	46	46.5	393	5 Q9V353	Q9v353 drosophila
14	46	46.5	438	2 Q845V3	Q845v3 burkholderi
15	45.5	46.0	596	16 Q8PHP2	Q8php2 xanthomonas
16	45	45.5	139	2 Q9XDG7	Q9xdg7 escherichia

17	45	45.5	271	10 Q851F8	Q851f8 oryza sativ
18	45	45.5	359	16 Q910G7	Q910g7 pseudomonas
19	45	45.5	692	11 Q8BR66	Q8br66 mus musculus
20	45	45.5	862	5 Q22354	Q22354 caenorhabdi
21	45	45.5	892	11 Q70254	Q70254 mus musculus
22	45	45.5	1151	5 Q44319	Q44319 anurida mar
23	45	45.5	1399	16 Q826G9	Q826g9 streptomyce
24	44	44.4	158	5 Q9VLR3	Q9vlr3 drosophila
25	44	44.4	256	5 Q8SUF1	Q8suf1 encephalito
26	44	44.4	263	13 Q8AVP9	Q8avp9 xenopus lae
27	44	44.4	296	16 Q8UEX7	Q8uex7 agrobacteri
28	44	44.4	376	16 Q9RYP2	Q9ryp2 deinococcus
29	44	44.4	478	16 Q9A9I9	Q9a9i9 caulobacter
30	44	44.4	514	16 Q8XW48	Q8xw48 raietonia s
31	44	44.4	590	2 Q93E16	Q93e16 rhizobium l
32	43.5	43.9	486	10 Q9SN94	Q9sn94 arabidopsis
33	43	43.4	126	4 Q9UHE2	Q9uhe2 homo sapien
34	43	43.4	126	17 Q8TJ72	Q8tj72 methanosarc
35	43	43.4	142	16 Q9AAB6	Q9aab6 caulobacter
36	43	43.4	203	16 Q9RTS1	Q9rts1 deinococcus
37	43	43.4	215	16 Q82CX3	Q82cx3 streptomyce
38	43	43.4	226	4 Q9Y5T7	Q9y5t7 homo sapien
39	43	43.4	280	17 Q27633	Q27633 methanobact
40	43	43.4	403	16 Q89IV3	Q89iv3 bradyrhizob
41	43	43.4	407	17 Q9HST2	Q9hst2 halobacteri
42	43	43.4	472	16 Q9KZK2	Q9kzk2 streptomyce
43	43	43.4	677	16 Q7UPA0	Q7upa0 rhodopirell
44	43	43.4	829	2 Q91UQ8	Q91uq8 rhizobium m
45	43	43.4	853	16 Q8ALU3	Q8alu3 bacteroides
46	43	43.4	1421	9 Q9FZU3	Q9fzu3 neisseria m
47	43	43.4	2454	3 Q9UV56	Q9uv56 emericella
48	43	43.4	2454	3 Q9UVP2	Q9uvp2 emericella
49	42.5	42.9	377	10 Q8LQW1	Q8lqw1 oryza sativ
50	42.5	42.9	521	12 Q66044	Q66044 caprine her
51	42.5	42.9	945	3 Q8X1V8	Q8x1v8 aspergillus
52	42.5	42.9	986	16 Q7UB87	Q7ube7 shigella fl
53	42.5	42.9	1128	17 Q51999	Q51999 halobacteri
54	42.5	42.9	1266	16 Q83Q01	Q83q01 shigella fl
55	42	42.4	105	12 Q9QBT8	Q9qbt8 potato roug
56	42	42.4	292	11 Q99KE2	Q99ke2 mus musculu
57	42	42.4	293	4 Q86SF8	Q86sf8 homo sapien
58	42	42.4	299	10 Q8W300	Q8w300 oryza sativ
59	42	42.4	306	6 Q77768	Q77768 cryotolagus
60	42	42.4	306	11 Q8C2G5	Q8c2g5 mus musculu
61	42	42.4	321	2 Q84BB3	Q84bb3 erwinia chr
62	42	42.4	330	16 Q8PRE1	Q8pre1 xanthomonas
63	42	42.4	333	16 Q89I32	Q89i32 bradyrhizob
64	42	42.4	339	16 Q89HK1	Q89hk1 bradyrhizob
65	42	42.4	341	16 Q7UHR5	Q7uhr5 rhodopirell
66	42	42.4	350	2 Q47506	Q47506 escherichia
67	42	42.4	350	2 Q83Y58	Q83y58 escherichia
68	42	42.4	358	5 Q22945	Q22945 caenorhabdi
69	42	42.4	360	16 Q7WN81	Q7wn81 bordetella
70	42	42.4	360	16 Q7W0V7	Q7w0v7 bordetella
71	42	42.4	375	17 Q9Y949	Q9y949 aeropyrum p
72	42	42.4	383	16 Q82CN1	Q82cn1 streptomyce
73	42	42.4	396	5 Q27702	Q27702 musca domes
74	42	42.4	409	16 Q8DV73	Q8dv73 streptococc
75	42	42.4	546	11 Q8BK66	Q8bk66 mus musculu
76	42	42.4	596	5 Q9V766	Q9v766 drosophila
77	42	42.4	639	5 Q9VXX5	Q9vxx5 drosophila
78	42	42.4	897	5 Q9VE88	Q9ve88 drosophila
79	42	42.4	922	5 Q21418	Q21418 caenorhabdi
80	42	42.4	1367	5 Q20120	Q20120 caenorhabdi
81	42	42.4	1826	4 Q8IXX4	Q8ixx4 homo sapien
82	41.5	41.9	132	16 Q89JZ0	Q89jz0 bradyrhizob
83	41.5	41.9	1266	16 Q8X9D5	Q8x9d5 escherichia
84	41.5	41.9	1266	16 Q8FD48	Q8fd48 escherichia
85	41	41.4	50	13 Q8JFC6	Q8jfc6 ficedula hy
86	41	41.4	51	13 Q8AJZ9	Q8ajz9 ficedula al
87	41	41.4	52	16 Q7UWA8	Q7uwa8 rhodopirell
88	41	41.4	54	13 Q8UF88	Q8ufe8 ficedula al
89	41	41.4	93	2 Q9KW24	Q9kw24 xanthomonas

90 41 41.4 127 16 Q8X6S6 escherichia
 91 41 41.4 127 16 Q8FCP2 escherichia
 92 41 41.4 127 16 Q8J1B5 shigella fl
 93 41 41.4 146 11 Q9D7V3 mus musculus
 94 41 41.4 171 11 Q9D7V3 mus musculus
 95 41 41.4 188 16 Q7UQX7 Q9XK7 rhodospirillum
 96 41 41.4 197 2 Q9X6H1 Q9X6H1 streptococcus
 97 41 41.4 206 9 Q38221 Q38221 bacteriophages
 98 41 41.4 229 16 Q8FJRS Q8FJRS escherichia
 99 41 41.4 238 5 Q17130 Q17130 babesia equ
 100 41 41.4 286 17 Q8TW67 Q8TW67 methanopyrus

ALIGNMENTS

RESULT 1

Q7XXP6 PRELIMINARY; PRT; 373 AA.
 AC Q7XXP6
 DT 01-OCT-2003 (TREMELrel. 25, Created)
 DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE Hypothetical protein.
 CS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Callus;
 RA Moriguchi K., Ito Y., Yamazaki Y., Kurata N.;
 RT "Finding of various plant nuclear proteins using yeast nuclear
 transport trap system - a proteomal approach."
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB110206; BAC78598.1; --
 KW Hypothetical protein.
 SQ SEQUENCE 373 AA; 37513 MW; 9B622770CFC6DA7 CRC64;

Query Match 52.5%; Score 52; DB 10; Length 373;
 Best Local Similarity 64.3%; Pred. No. 8.7;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 7 LALSGSDGRVSGC 20
 :||:|||||:
 DB 254 VALAGSDGRVGGC 267

RESULT 2

Q7XM00 PRELIMINARY; PRT; 379 AA.
 AC Q7XM00
 DT 01-OCT-2003 (TREMELrel. 25, Created)
 DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE OSJNBA0086006.9 protein.
 GN OSJNBA0086006.9
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Han B., Feng Q., Huang Y.C., Li Y., Zhu J.J., Zhao Q., Hu X.,
 RA Liu Y.L., Mu J., Yu Z., Chen L., Fan D.L., Weng Q.J., Zhang L.,
 RA Lu Y.Q., Yu S.L., Liu X.H., Lu T.T., Zhang Y.J., Lu Y., Li C., Li T.,
 RA Zhang Y., Hu H., Jia P.X., Qian Y.M., Ying K., Zhou B., Chen Z.H.,
 RA Hao P., Zhang L., Wu M., Zhang R.Q., Guan J.P., Fu G., Wang S.Y.,
 RA Ren S.X., Lv G., Lin W., Zhu G.F., Tu Y.F., Jia J., Yin H.F.,
 RA Zhang Y., Cai Z., Chen J., Kang H., Chen X.Y., Shao C.Y., Sun Y.,
 RA Hu Q.P., Zhang X.L., Zhang W., Wang L.J., Ding C.W., Sheng H.H.,

Query Match 52.5%; Score 52; DB 10; Length 405;
 Best Local Similarity 71.4%; Pred. No. 9.5;
 Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 7 LALSGSDGRVSGC 20
 :||:|||||:
 DB 272 VALAGSDGRVGGC 285

RESULT 4

Q7X976 PRELIMINARY; PRT; 405 AA.
 AC Q7X976
 DT 01-OCT-2003 (TREMELrel. 25, Created)
 DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE Putative AT-hook DNA-binding protein.
 GN OSJNBA003019.1
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=39947;

RA Gu J.L., Chen S.T., Ni L., Zhu F.H., Hong G.F.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL662981; CAE04861.1; -- FC493F8D769E4670 CRC64;
 SQ SEQUENCE 379 AA; 38241 MW; 774994FD12D49816 CRC64;
 Query Match 52.5%; Score 52; DB 10; Length 379;
 Best Local Similarity 64.3%; Pred. No. 8.8;
 Matches 9; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 7 LALSGSDGRVSGC 20
 :||:|||||:
 DB 254 VALAGSDGRVGGC 267

RESULT 3

Q9AYM0 PRELIMINARY; PRT; 405 AA.
 AC Q9AYM0
 DT 01-JUN-2001 (TREMELrel. 17, Created)
 DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)
 DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 DE Putative AT-hook DNA-binding protein.
 GN OSJNBA003019.1
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Buell C.R., Yuan Q., Moffat K.S., Hill J.N., Burr P.C., Heiao J.,
 RA Ziemann V., Pai G., Bowman C.L., Fujii C.Y., VanAken S.E.,
 RA Bowman C.L., Craven B., Utterback T.R., Khalak H., Feldblyum T.V.,
 RA Quackenbush J., White O., Salzberg S.L., Fraser C.M.;
 RT "Oryza sativa chromosome 10 BAC OSJNBA003019 genomic sequence."
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC060755; AAK00433.1; --
 DR Gramene; Q9AYM0; --
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR000637; AT hook.
 DR InterPro; IPR005175; DUF296.
 DR Pfam; PF02178; AT hook; 2.
 DR Pfam; PF03479; DUF296; 1.
 DR SMART; SM00384; AT_hook; 2.
 KW DNA-binding.
 SQ SEQUENCE 405 AA; 41399 MW; 274994FD12D49816 CRC64;

Query Match 52.5%; Score 52; DB 10; Length 405;
 Best Local Similarity 71.4%; Pred. No. 9.5;
 Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 7 LALSGSDGRVSGC 20
 :||:|||||:
 DB 272 VALAGSDGRVGGC 285


```

RN RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA The Rice Chromosome 10 Sequencing Consortium;
RT "In-depth view of structure, activity, and evolution of rice
RL chromosome 10.";
RL Science 300:1566-1569(2003).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Buell C.R., Wang R.A., McCombie W.R., Messing J., Yuan Q.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB017121; AAP55117.1; -.
KW DNA-binding.
SQ SEQUENCE 405 AA; 41399 MW; 274994FD12D49816 CRC64;

Query Match 52.5%; Score 52; DB 10; Length 405;
Best Local Similarity 71.4%; Pred. No. 9.5;
Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 7 LALSGDGRVSGGC 20
DB 272 VALAGSDGRVLCGC 285
:|||||
[1]
[2]

RESULT 5
Q8EHA4 PRELIMINARY; PRT; 1482 AA.
AC Q8EHA4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Glutamate synthase, large subunit.
GN GLTB OR SOI325.
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Alteromonadaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MR-1;
RX MEDLINE=22297686; PubMed=1236813;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N., Methe B., Clayton R.A.,
RA Meyer T., Tsapin A., Scott J., Beanan M., Brinkac L., Daugherty S.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
RA Mueller J., Khouri H., Gill J., Utterback T.R., McDonald L.A.,
RA Feldblyum T.V., Smith H.O., Venter J.C., Neallson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis.";
RL Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL; AB015576; AAN54390.1; -.
DR TIGR; SO1325; -.
DR GO; GO:0015930; F:glutamate synthase activity; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:006537; P:glutamate biosynthesis; IEA.
DR GO; GO:0006807; P:nitrogen metabolism; IEA.
DR InterPro; IPR002198; ADH_short.
DR InterPro; IPR002489; DUF14.
DR InterPro; IPR003009; FMN enzyme.
DR InterPro; IPR002932; Glu synthase.
DR InterPro; IPR006982; Glu synth centr.
DR InterPro; IPR006981; Glu synth NTN.
DR Pfam; PF01645; Glu_synthase; 1.
DR Pfam; PF04897; Glu synth NTN; 1.
DR Pfam; PF04898; Glu syn centr; 1.
DR Pfam; PF01493; GXGKG; 1.
DR PROSITE; PS00061; ADH_SHORT; 1.
KW Complete proteome.
SQ SEQUENCE 1482 AA; 163011 MW; 10FDB14B658BF474 CRC64;

Query Match 50.0%; Score 49.5; DB 16; Length 1482;
Best Local Similarity 41.4%; Pred. No. 94;
Matches 12; Conservative 1; Mismatches 5; Indels 11; Gaps 1;

QY 3 RAVRLALSG-----SDGRVSGGC 20
DB 28 RIVRTAIRGLDRMKHRRGIASDGRVGDGC 56
:|||||
[1]
[2]

RESULT 6
O28180 PRELIMINARY; PRT; 575 AA.
AC O28180;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Acetolactate synthase, large subunit (ILV5-4).
GN AF2100.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey B.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC -1- SIMILARITY: WITH OTHER ENZYMES WHICH REQUIRE TPP.
DR EMBL; AB008958; AAB89146.1; -.
DR FIR; D69512; D69512.
DR HSSP; P07342; 1JSC.
DR TIGR; AF2100; -.
DR InterPro; IPR000399; Pyruvate decarb.
DR Pfam; PF00205; TPP_enzymes; 1.
DR Pfam; PF02775; TPP_enzymes; 1.
DR Pfam; PF02776; TPP_enzymes; 1.
DR PROSITE; PS00187; TPP_ENZYMES; 1.
KW Flavoprotein; Hypothetical protein; Thiamine pyrophosphate;
KW Complete proteome.
SQ SEQUENCE 575 AA; 63000 MW; B121765C66EC3E8A CRC64;

Query Match 49.5%; Score 49; DB 17; Length 575;
Best Local Similarity 62.5%; Pred. No. 41;
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 LQRAVRLALSGSDGRV 16
DB 147 VQRAIRIALSGRPGPV 162
:|||||
[1]
[2]

RESULT 7
Q60544 PRELIMINARY; PRT; 1865 AA.
AC Q60544;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE CCG1.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OX Mesocricetus.
RN NCBI_TaxID=10036;
RP SEQUENCE FROM N.A.
RX MEDLINE=94215915; PubMed=8163200;
RA Hayashida T., Sekiguchi T., Noguchi E., Sunamoto H., Ohba T.,
RA Nishimoto T.;
RT "The CGG17AFI1250 gene is mutated in thermosensitive G1 mutants of
RT the BHK21 cell line derived from goldenhamster.";
RL Gene 141:267-270(1994).
DR EMBL; D26114; BAA05110.1; -.
DR PIR; I48155; I48155.
DR InterPro; IPR001487; Bromodomain.
DR PRINTS; PR00503; bromodomain; 2.
DR SMART; SM00297; BROMO; 2.
DR PROSITE; PS00633; BROMODOMAIN_1; 2.
DR PROSITE; PS0014; BROMODOMAIN_2; 2.
SQ SEQUENCE 1865 AA; 211866 MW; A81614946C0CF24 CRC64;

Query Match 48.5%; Score 48; DB 11; Length 1865;
Best Local Similarity 62.5%; Pred. No. 2.1e-02;
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 AVRLALSGSDGRVGS 19
Db ::::|||||
1735 AIQLSESGSDSDVGS 1750

RESULT 8
Q8U4U3 PRELIMINARY; PRT; 114 AA.
AC Q8U4U3;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Putative nitrogen regulatory protein GlnK1.
GN GlnK1.
OS Methanosarcina mazei (Methanosarcina frisia).
OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2209;
RP SEQUENCE FROM N.A.
RX STRAIN=Goel;
RX MEDLINE=21666136; PubMed=11807063;
RA Ehlers C., Grabbe R., Veit K., Schmitz R.A.;
RT "Characterization of GlnK(1) from Methanosarcina mazei Strain Goel:
RT Complementation of an Escherichia coli glnK Mutant Strain by
RT GlnK(1).";
RL J. Bacteriol. 184:1028-1040(2002).
DR EMBL; AF367242; AAL74059.1; -.
DR GO; GO:0030234; F:enzyme regulator activity; IEA.
DR GO; GO:0006808; P:regulation of nitrogen utilization; IEA.
DR InterPro; IPR002187; PII_glnB.
DR Pfam; PF00543; P-II; 1.
DR PRINTS; PR00340; PIIGLNB.
DR ProDom; PD001194; PII_glnB; 1.
SQ SEQUENCE 114 AA; 12753 MW; FD1BAFA84C4B7261 CRC64;

Query Match 47.5%; Score 47; DB 1; Length 114;
Best Local Similarity 56.2%; Pred. No. 15;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 4 AVRLALSGSDGRVGS 19
Db ::::|||||
76 AIKTGAKGSDGVGS 91

RESULT 9
Q8PYW7 PRELIMINARY; PRT; 117 AA.
ID Q8PYW7
AC Q8RMH1 PRELIMINARY; PRT; 325 AA.
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE NifR3-like protein.
GN NifR3.
OS Acetobacter diazotrophicus (Gluconacetobacter diazotrophicus).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
OC Acetobacteraceae; Gluconacetobacter.
OX NCBI_TaxID=33996;
RP SEQUENCE FROM N.A.
RA Perlova O., Nawroth R., Baumgarth B., Agnieszka S., Meletzus D.;
RT "Molecular analysis of the chromosomal nifBC region of Acetobacter
RT diazotrophicus.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF494454; AAM15930.1; -.
DR InterPro; IPR003009; FMN enzyme.
DR InterPro; IPR004652; NifR3_YhdG.
DR InterPro; IPR001269; UPF0034.
DR Pfam; PF01207; Duf; 1.
DR TIGRFAMs; TIGR00737; nifR3_YhdG; 1.
SQ SEQUENCE 325 AA; 34791 MW; E6CAE79BFE5F8256 CRC64;

Query Match 47.0%; Score 46.5; DB 2; Length 325;
Best Local Similarity 66.7%; Pred. No. 55;
Matches 10; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

QY 7 LALSGSDG-RVGS 20

```

```

|||||:|:|:|:|:|:|
195 LALSGADGVMIGRC 209

RESULT 11
Q22AC6 PRELIMINARY; PRT; 352 AA.
ID Q92AC6
AC Q92AC6
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Hypothetical protein lin1996.
GN Lin1996.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / Serovar 6a;
RX MEDLINE=21537279; PubMed=11679669;
RA Baquero F., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Glaeser P., Berche P., Bloecker H., Brandt P., Chakraborty T.,
RA Chabrit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Krest J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehlend J., Cossart P.;
RT "Comparative Genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL596170; CAC97226.1; -.
DR PIR; AB1682; AB1682.
DR Listlist; LIN01996; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001223; Glyco_hydro_18.
DR InterPro; IPR001579; Glyco_hydro_18AS.
DR Pfam; PF00704; Glyco_hydro_18; 1.
DR ProDom; PD000471; Glyco_hydro_18; 1.
DR PROSITE; PS01095; CHITINASE 18; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 352 AA; 37949 MW; 771A0DC781C11359 CRC64;

Query Match 46.5%; Score 46; DB 16; Length 352;
Best Local Similarity 71.4%; Pred. No. 72;
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 RAVRLALSGSGGRV 16
DB 119 RAVLLALGGADGHV 132

RESULT 12
Q9V353 PRELIMINARY; PRT; 393 AA.
ID Q9V353
AC Q9V353
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE CG8719 protein.
GN ODC2 OR CG8719.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadenat E., Center A., Chandra I.,
RA Cherty J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Hartis N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

```

```

RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Krest J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehlend J., Cossart P.;
RT "Comparative Genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL591981; CAC99961.1; -.
DR PIR; AC1310; AC1310.
DR Listlist; LMO01883; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001223; Glyco_hydro_18.
DR InterPro; IPR001579; Glyco_hydro_18AS.
DR Pfam; PF00704; Glyco_hydro_18; 1.
DR ProDom; PD000471; Glyco_hydro_18; 1.
DR PROSITE; PS01095; CHITINASE 18; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 352 AA; 37920 MW; CEE13D93A4858A88 CRC64;

Query Match 46.5%; Score 46; DB 16; Length 352;
Best Local Similarity 71.4%; Pred. No. 72;
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 RAVRLALSGSGGRV 16
DB 119 RAVLLALGGADGHV 132

RESULT 13
Q9V353 PRELIMINARY; PRT; 393 AA.
ID Q9V353
AC Q9V353
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE CG8719 protein.
GN ODC2 OR CG8719.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadenat E., Center A., Chandra I.,
RA Cherty J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Hartis N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

```

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry J., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupeki M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 [2]
 RP SEQUENCE FROM N.A.
 RA Celinker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banton J., An H., Baldwin D., Banton J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dreenek D., Farfan D.,
 RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacleb J., Paragay V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RA "Sequencing of *Drosophila melanogaster* genome.";
 RT Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases.
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celinker S.E.,
 RA Clump M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RA "Annotation of *Drosophila melanogaster* genome.";
 RT Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases.
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celinker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBSJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBSJ databases.
 DR EMBL; AF003839; RAF59149.2; -;
 DR HSP; P07805; 1F3T.
 DR FlyBase; FBgn0013308; Odc2.
 DR GO; GO:0003824; F-catalytic activity; IEA.
 DR InterPro; IPR000183; Decarboxylase2.
 DR InterPro; IPR009006; Racem decarbox-C.
 DR Pfam; PF02784; Orn Arg dec N; 1.
 DR Pfam; PF00278; Orn Arg dec; 1.
 DR PRINTS; PR01179; ODABCBXKAGE.
 DR PROSITE; PS00878; ODR DC 2.1; 1.
 DR PROSITE; PS00879; ODR DC 2.2; 1.
 DR PROSITE; PS00879; ODR DC 2.3; 1.
 DR PROSITE; PS00879; ODR DC 2.4; 1.
 DR PROSITE; PS00879; ODR DC 2.5; 1.
 DR PROSITE; PS00879; ODR DC 2.6; 1.
 DR PROSITE; PS00879; ODR DC 2.7; 1.
 DR PROSITE; PS00879; ODR DC 2.8; 1.
 DR PROSITE; PS00879; ODR DC 2.9; 1.
 DR PROSITE; PS00879; ODR DC 2.10; 1.
 DR PROSITE; PS00879; ODR DC 2.11; 1.
 DR PROSITE; PS00879; ODR DC 2.12; 1.
 DR PROSITE; PS00879; ODR DC 2.13; 1.
 DR PROSITE; PS00879; ODR DC 2.14; 1.
 DR PROSITE; PS00879; ODR DC 2.15; 1.
 DR PROSITE; PS00879; ODR DC 2.16; 1.
 DR PROSITE; PS00879; ODR DC 2.17; 1.
 DR PROSITE; PS00879; ODR DC 2.18; 1.
 DR PROSITE; PS00879; ODR DC 2.19; 1.
 DR PROSITE; PS00879; ODR DC 2.20; 1.
 DR PROSITE; PS00879; ODR DC 2.21; 1.
 DR PROSITE; PS00879; ODR DC 2.22; 1.
 DR PROSITE; PS00879; ODR DC 2.23; 1.
 DR PROSITE; PS00879; ODR DC 2.24; 1.
 DR PROSITE; PS00879; ODR DC 2.25; 1.
 DR PROSITE; PS00879; ODR DC 2.26; 1.
 DR PROSITE; PS00879; ODR DC 2.27; 1.
 DR PROSITE; PS00879; ODR DC 2.28; 1.
 DR PROSITE; PS00879; ODR DC 2.29; 1.
 DR PROSITE; PS00879; ODR DC 2.30; 1.
 DR PROSITE; PS00879; ODR DC 2.31; 1.
 DR PROSITE; PS00879; ODR DC 2.32; 1.
 DR PROSITE; PS00879; ODR DC 2.33; 1.
 DR PROSITE; PS00879; ODR DC 2.34; 1.
 DR PROSITE; PS00879; ODR DC 2.35; 1.
 DR PROSITE; PS00879; ODR DC 2.36; 1.
 DR PROSITE; PS00879; ODR DC 2.37; 1.
 DR PROSITE; PS00879; ODR DC 2.38; 1.
 DR PROSITE; PS00879; ODR DC 2.39; 1.
 DR PROSITE; PS00879; ODR DC 2.40; 1.
 DR PROSITE; PS00879; ODR DC 2.41; 1.
 DR PROSITE; PS00879; ODR DC 2.42; 1.
 DR PROSITE; PS00879; ODR DC 2.43; 1.
 DR PROSITE; PS00879; ODR DC 2.44; 1.
 DR PROSITE; PS00879; ODR DC 2.45; 1.
 DR PROSITE; PS00879; ODR DC 2.46; 1.
 DR PROSITE; PS00879; ODR DC 2.47; 1.
 DR PROSITE; PS00879; ODR DC 2.48; 1.
 DR PROSITE; PS00879; ODR DC 2.49; 1.
 DR PROSITE; PS00879; ODR DC 2.50; 1.
 DR PROSITE; PS00879; ODR DC 2.51; 1.
 DR PROSITE; PS00879; ODR DC 2.52; 1.
 DR PROSITE; PS00879; ODR DC 2.53; 1.
 DR PROSITE; PS00879; ODR DC 2.54; 1.
 DR PROSITE; PS00879; ODR DC 2.55; 1.
 DR PROSITE; PS00879; ODR DC 2.56; 1.
 DR PROSITE; PS00879; ODR DC 2.57; 1.
 DR PROSITE; PS00879; ODR DC 2.58; 1.
 DR PROSITE; PS00879; ODR DC 2.59; 1.
 DR PROSITE; PS00879; ODR DC 2.60; 1.
 DR PROSITE; PS00879; ODR DC 2.61; 1.
 DR PROSITE; PS00879; ODR DC 2.62; 1.
 DR PROSITE; PS00879; ODR DC 2.63; 1.
 DR PROSITE; PS00879; ODR DC 2.64; 1.
 DR PROSITE; PS00879; ODR DC 2.65; 1.
 DR PROSITE; PS00879; ODR DC 2.66; 1.
 DR PROSITE; PS00879; ODR DC 2.67; 1.
 DR PROSITE; PS00879; ODR DC 2.68; 1.
 DR PROSITE; PS00879; ODR DC 2.69; 1.
 DR PROSITE; PS00879; ODR DC 2.70; 1.
 DR PROSITE; PS00879; ODR DC 2.71; 1.
 DR PROSITE; PS00879; ODR DC 2.72; 1.
 DR PROSITE; PS00879; ODR DC 2.73; 1.
 DR PROSITE; PS00879; ODR DC 2.74; 1.
 DR PROSITE; PS00879; ODR DC 2.75; 1.
 DR PROSITE; PS00879; ODR DC 2.76; 1.
 DR PROSITE; PS00879; ODR DC 2.77; 1.
 DR PROSITE; PS00879; ODR DC 2.78; 1.
 DR PROSITE; PS00879; ODR DC 2.79; 1.
 DR PROSITE; PS00879; ODR DC 2.80; 1.
 DR PROSITE; PS00879; ODR DC 2.81; 1.
 DR PROSITE; PS00879; ODR DC 2.82; 1.
 DR PROSITE; PS00879; ODR DC 2.83; 1.
 DR PROSITE; PS00879; ODR DC 2.84; 1.
 DR PROSITE; PS00879; ODR DC 2.85; 1.
 DR PROSITE; PS00879; ODR DC 2.86; 1.
 DR PROSITE; PS00879; ODR DC 2.87; 1.
 DR PROSITE; PS00879; ODR DC 2.88; 1.
 DR PROSITE; PS00879; ODR DC 2.89; 1.
 DR PROSITE; PS00879; ODR DC 2.90; 1.
 DR PROSITE; PS00879; ODR DC 2.91; 1.
 DR PROSITE; PS00879; ODR DC 2.92; 1.
 DR PROSITE; PS00879; ODR DC 2.93; 1.
 DR PROSITE; PS00879; ODR DC 2.94; 1.
 DR PROSITE; PS00879; ODR DC 2.95; 1.
 DR PROSITE; PS00879; ODR DC 2.96; 1.
 DR PROSITE; PS00879; ODR DC 2.97; 1.
 DR PROSITE; PS00879; ODR DC 2.98; 1.
 DR PROSITE; PS00879; ODR DC 2.99; 1.
 DR PROSITE; PS00879; ODR DC 3.00; 1.
 DR PROSITE; PS00879; ODR DC 3.01; 1.
 DR PROSITE; PS00879; ODR DC 3.02; 1.
 DR PROSITE; PS00879; ODR DC 3.03; 1.
 DR PROSITE; PS00879; ODR DC 3.04; 1.
 DR PROSITE; PS00879; ODR DC 3.05; 1.
 DR PROSITE; PS00879; ODR DC 3.06; 1.
 DR PROSITE; PS00879; ODR DC 3.07; 1.
 DR PROSITE; PS00879; ODR DC 3.08; 1.
 DR PROSITE; PS00879; ODR DC 3.09; 1.
 DR PROSITE; PS00879; ODR DC 3.10; 1.
 DR PROSITE; PS00879; ODR DC 3.11; 1.
 DR PROSITE; PS00879; ODR DC 3.12; 1.
 DR PROSITE; PS00879; ODR DC 3.13; 1.
 DR PROSITE; PS00879; ODR DC 3.14; 1.
 DR PROSITE; PS00879; ODR DC 3.15; 1.
 DR PROSITE; PS00879; ODR DC 3.16; 1.
 DR PROSITE; PS00879; ODR DC 3.17; 1.
 DR PROSITE; PS00879; ODR DC 3.18; 1.
 DR PROSITE; PS00879; ODR DC 3.19; 1.
 DR PROSITE; PS00879; ODR DC 3.20; 1.
 DR PROSITE; PS00879; ODR DC 3.21; 1.
 DR PROSITE; PS00879; ODR DC 3.22; 1.
 DR PROSITE; PS00879; ODR DC 3.23; 1.
 DR PROSITE; PS00879; ODR DC 3.24; 1.
 DR PROSITE; PS00879; ODR DC 3.25; 1.
 DR PROSITE; PS00879; ODR DC 3.26; 1.
 DR PROSITE; PS00879; ODR DC 3.27; 1.
 DR PROSITE; PS00879; ODR DC 3.28; 1.
 DR PROSITE; PS00879; ODR DC 3.29; 1.
 DR PROSITE; PS00879; ODR DC 3.30; 1.
 DR PROSITE; PS00879; ODR DC 3.31; 1.
 DR PROSITE; PS00879; ODR DC 3.32; 1.
 DR PROSITE; PS00879; ODR DC 3.33; 1.
 DR PROSITE; PS00879; ODR DC 3.34; 1.
 DR PROSITE; PS00879; ODR DC 3.35; 1.
 DR PROSITE; PS00879; ODR DC 3.36; 1.
 DR PROSITE; PS00879; ODR DC 3.37; 1.
 DR PROSITE; PS00879; ODR DC 3.38; 1.
 DR PROSITE; PS00879; ODR DC 3.39; 1.
 DR PROSITE; PS00879; ODR DC 3.40; 1.
 DR PROSITE; PS00879; ODR DC 3.41; 1.
 DR PROSITE; PS00879; ODR DC 3.42; 1.
 DR PROSITE; PS00879; ODR DC 3.43; 1.
 DR PROSITE; PS00879; ODR DC 3.44; 1.
 DR PROSITE; PS00879; ODR DC 3.45; 1.
 DR PROSITE; PS00879; ODR DC 3.46; 1.
 DR PROSITE; PS00879; ODR DC 3.47; 1.
 DR PROSITE; PS00879; ODR DC 3.48; 1.
 DR PROSITE; PS00879; ODR DC 3.49; 1.
 DR PROSITE; PS00879; ODR DC 3.50; 1.
 DR PROSITE; PS00879; ODR DC 3.51; 1.
 DR PROSITE; PS00879; ODR DC 3.52; 1.
 DR PROSITE; PS00879; ODR DC 3.53; 1.
 DR PROSITE; PS00879; ODR DC 3.54; 1.
 DR PROSITE; PS00879; ODR DC 3.55; 1.
 DR PROSITE; PS00879; ODR DC 3.56; 1.
 DR PROSITE; PS00879; ODR DC 3.57; 1.
 DR PROSITE; PS00879; ODR DC 3.58; 1.
 DR PROSITE; PS00879; ODR DC 3.59; 1.
 DR PROSITE; PS00879; ODR DC 3.60; 1.
 DR PROSITE; PS00879; ODR DC 3.61; 1.
 DR PROSITE; PS00879; ODR DC 3.62; 1.
 DR PROSITE; PS00879; ODR DC 3.63; 1.
 DR PROSITE; PS00879; ODR DC 3.64; 1.
 DR PROSITE; PS00879; ODR DC 3.65; 1.
 DR PROSITE; PS00879; ODR DC 3.66; 1.
 DR PROSITE; PS00879; ODR DC 3.67; 1.
 DR PROSITE; PS00879; ODR DC 3.68; 1.
 DR PROSITE; PS00879; ODR DC 3.69; 1.
 DR PROSITE; PS00879; ODR DC 3.70; 1.
 DR PROSITE; PS00879; ODR DC 3.71; 1.
 DR PROSITE; PS00879; ODR DC 3.72; 1.
 DR PROSITE; PS00879; ODR DC 3.73; 1.
 DR PROSITE; PS00879; ODR DC 3.74; 1.
 DR PROSITE; PS00879; ODR DC 3.75; 1.
 DR PROSITE; PS00879; ODR DC 3.76; 1.
 DR PROSITE; PS00879; ODR DC 3.77; 1.
 DR PROSITE; PS00879; ODR DC 3.78; 1.
 DR PROSITE; PS00879; ODR DC 3.79; 1.
 DR PROSITE; PS00879; ODR DC 3.80; 1.
 DR PROSITE; PS00879; ODR DC 3.81; 1.
 DR PROSITE; PS00879; ODR DC 3.82; 1.
 DR PROSITE; PS00879; ODR DC 3.83; 1.
 DR PROSITE; PS00879; ODR DC 3.84; 1.
 DR PROSITE; PS00879; ODR DC 3.85; 1.
 DR PROSITE; PS00879; ODR DC 3.86; 1.
 DR PROSITE; PS00879; ODR DC 3.87; 1.
 DR PROSITE; PS00879; ODR DC 3.88; 1.
 DR PROSITE; PS00879; ODR DC 3.89; 1.
 DR PROSITE; PS00879; ODR DC 3.90; 1.
 DR PROSITE; PS00879; ODR DC 3.91; 1.
 DR PROSITE; PS00879; ODR DC 3.92; 1.
 DR PROSITE; PS00879; ODR DC 3.93; 1.
 DR PROSITE; PS00879; ODR DC 3.94; 1.
 DR PROSITE; PS00879; ODR DC 3.95; 1.
 DR PROSITE; PS00879; ODR DC 3.96; 1.
 DR PROSITE; PS00879; ODR DC 3.97; 1.
 DR PROSITE; PS00879; ODR DC 3.98; 1.
 DR PROSITE; PS00879; ODR DC 3.99; 1.
 DR PROSITE; PS00879; ODR DC 4.00; 1.
 DR PROSITE; PS00879; ODR DC 4.01; 1.
 DR PROSITE; PS00879; ODR DC 4.02; 1.
 DR PROSITE; PS00879; ODR DC 4.03; 1.
 DR PROSITE; PS00879; ODR DC 4.04; 1.
 DR PROSITE; PS00879; ODR DC 4.05; 1.
 DR PROSITE; PS00879; ODR DC 4.06; 1.
 DR PROSITE; PS00879; ODR DC 4.07; 1.
 DR PROSITE; PS00879; ODR DC 4.08; 1.
 DR PROSITE; PS00879; ODR DC 4.09; 1.
 DR PROSITE; PS00879; ODR DC 4.10; 1.
 DR PROSITE; PS00879; ODR DC 4.11; 1.
 DR PROSITE; PS00879; ODR DC 4.12; 1.
 DR PROSITE; PS00879; ODR DC 4.13; 1.
 DR PROSITE; PS00879; ODR DC 4.14; 1.
 DR PROSITE; PS00879; ODR DC 4.15; 1.
 DR PROSITE; PS00879; ODR DC 4.16; 1.
 DR PROSITE; PS00879; ODR DC 4.17; 1.
 DR PROSITE; PS00879; ODR DC 4.18; 1.
 DR PROSITE; PS00879; ODR DC 4.19; 1.
 DR PROSITE; PS00879; ODR DC 4.20; 1.
 DR PROSITE; PS00879; ODR DC 4.21; 1.
 DR PROSITE; PS00879; ODR DC 4.22; 1.
 DR PROSITE; PS00879; ODR DC 4.23; 1.
 DR PROSITE; PS00879; ODR DC 4.24; 1.
 DR PROSITE; PS00879; ODR DC 4.25; 1.
 DR PROSITE; PS00879; ODR DC 4.26; 1.
 DR PROSITE; PS00879; ODR DC 4.27; 1.
 DR PROSITE; PS00879; ODR DC 4.28; 1.
 DR PROSITE; PS00879; ODR DC 4.29; 1.
 DR PROSITE; PS00879; ODR DC 4.30; 1.
 DR PROSITE; PS00879; ODR DC 4.31; 1.
 DR PROSITE; PS00879; ODR DC 4.32; 1.
 DR PROSITE; PS00879; ODR DC 4.33; 1.
 DR PROSITE; PS00879; ODR DC 4.34; 1.
 DR PROSITE; PS00879; ODR DC 4.35; 1.
 DR PROSITE; PS00879; ODR DC 4.36; 1.
 DR PROSITE; PS00879; ODR DC 4.37; 1.
 DR PROSITE; PS00879; ODR DC 4.38; 1.
 DR PROSITE; PS00879; ODR DC 4.39; 1.
 DR PROSITE; PS00879; ODR DC 4.40; 1.
 DR PROSITE; PS00879; ODR DC 4.41; 1.
 DR PROSITE; PS00879; ODR DC 4.42; 1.
 DR PROSITE; PS00879; ODR DC 4.43; 1.
 DR PROSITE; PS00879; ODR DC 4.44; 1.
 DR PROSITE; PS00879; ODR DC 4.45; 1.
 DR PROSITE; PS00879; ODR DC 4.46; 1.
 DR PROSITE; PS00879; ODR DC 4.47; 1.
 DR PROSITE; PS00879; ODR DC 4.48; 1.
 DR PROSITE; PS00879; ODR DC 4.49; 1.
 DR PROSITE; PS00879; ODR DC 4.50; 1.
 DR PROSITE; PS00879; ODR DC 4.51; 1.
 DR PROSITE; PS00879; ODR DC 4.52; 1.
 DR PROSITE; PS00879; ODR DC 4.53; 1.
 DR PROSITE; PS00879; ODR DC 4.54; 1.
 DR PROSITE; PS00879; ODR DC 4.55; 1.
 DR PROSITE; PS00879; ODR DC 4.56; 1.
 DR PROSITE; PS00879; ODR DC 4.57; 1.
 DR PROSITE; PS00879; ODR DC 4.58; 1.
 DR PROSITE; PS00879; ODR DC 4.59; 1.
 DR PROSITE; PS00879; ODR DC 4.60; 1.
 DR PROSITE; PS00879; ODR DC 4.61; 1.
 DR PROSITE; PS00879; ODR DC 4.62; 1.
 DR PROSITE; PS00879; ODR DC 4.63; 1.
 DR PROSITE; PS00879; ODR DC 4.64; 1.
 DR PROSITE; PS00879; ODR DC 4.65; 1.
 DR PROSITE; PS00879; ODR DC 4.66; 1.
 DR PROSITE; PS00879; ODR DC 4.67; 1.
 DR PROSITE; PS00879; ODR DC 4.68; 1.
 DR PROSITE; PS00879; ODR DC 4.69; 1.
 DR PROSITE; PS00879; ODR DC 4.70; 1.
 DR PROSITE; PS00879; ODR DC 4.71; 1.
 DR PROSITE; PS00879; ODR DC 4.72; 1.
 DR PROSITE; PS00879; ODR DC 4.73; 1.
 DR PROSITE; PS00879; ODR DC 4.74; 1.
 DR PROSITE; PS00879; ODR DC 4.75; 1.
 DR PROSITE; PS00879; ODR DC 4.76; 1.
 DR PROSITE; PS00879; ODR DC 4.77; 1.
 DR PROSITE; PS00879; ODR DC 4.78; 1.
 DR PROSITE; PS00879; ODR DC 4.79; 1.
 DR PROSITE; PS00879; ODR DC 4.80; 1.
 DR PROSITE; PS00879; ODR DC 4.81; 1.
 DR PROSITE; PS00879; ODR DC 4.82; 1.
 DR PROSITE; PS00879; ODR DC 4.83; 1.
 DR PROSITE; PS00879; ODR DC 4.84; 1.
 DR PROSITE; PS00879; ODR DC 4.85; 1.
 DR PROSITE; PS00879; ODR DC 4.86; 1.
 DR PROSITE; PS00879; ODR DC 4.87; 1.
 DR PROSITE; PS00879; ODR DC 4.88; 1.
 DR PROSITE; PS00879; ODR DC 4.89; 1.
 DR PROSITE; PS00879; ODR DC 4.90; 1.
 DR PROSITE; PS00879; ODR DC 4.91; 1.
 DR PROSITE; PS00879; ODR DC 4.92; 1.
 DR PROSITE; PS00879; ODR DC 4.93; 1.
 DR PROSITE; PS00879; ODR DC 4.94; 1.
 DR PROSITE; PS00879; ODR DC 4.95; 1.
 DR PROSITE; PS00879; ODR DC 4.96; 1.
 DR PROSITE; PS00879; ODR DC 4.97; 1.
 DR PROSITE; PS00879; ODR DC 4.98; 1.
 DR PROSITE; PS00879; ODR DC 4.99; 1.
 DR PROSITE; PS00879; ODR DC 5.00; 1.
 DR PROSITE; PS00879; ODR DC 5.01; 1.
 DR PROSITE; PS00879; ODR DC 5.02; 1.
 DR PROSITE; PS00879; ODR DC 5.03; 1.
 DR PROSITE; PS00879; ODR DC 5.04; 1.
 DR PROSITE; PS00879; ODR DC 5.05; 1.
 DR PROSITE; PS00879; ODR DC 5.06; 1.
 DR PROSITE; PS00879; ODR DC 5.07; 1.
 DR PROSITE; PS00879; ODR DC 5.08; 1.
 DR PROSITE; PS00879; ODR DC 5.09; 1.
 DR PROSITE; PS00879; ODR DC 5.10; 1.
 DR PROSITE; PS00879; ODR DC 5.11; 1.
 DR PROSITE; PS00879; ODR DC 5.12; 1.
 DR PROSITE; PS00879; ODR DC 5.13; 1.
 DR PROSITE; PS00879; ODR DC 5.14; 1.
 DR PROSITE; PS00879; ODR DC 5.15; 1.
 DR PROSITE; PS00879; ODR DC 5.16; 1.
 DR PROSITE; PS00879; ODR DC 5.17; 1.
 DR PROSITE; PS00879; ODR DC 5.18; 1.
 DR PROSITE; PS00879; ODR DC 5.19; 1.
 DR PROSITE; PS00879; ODR DC 5.20; 1.
 DR PROSITE; PS00879; ODR DC 5.21; 1.
 DR PROSITE; PS00879; ODR DC 5.22; 1.
 DR PROSITE; PS00879; ODR DC 5.23; 1.
 DR PROSITE; PS00879; ODR DC 5.24; 1.
 DR PROSITE; PS00879; ODR DC 5.25; 1.
 DR PROSITE; PS00879; ODR DC 5.26; 1.
 DR PROSITE; PS00879; ODR DC 5.27; 1.
 DR PROSITE; PS00879; ODR DC 5.28; 1.
 DR PROSITE; PS00879; ODR DC 5.29; 1.
 DR PROSITE; PS00879; ODR DC 5.30; 1.
 DR PROSITE; PS00879; ODR DC 5.31; 1.
 DR PROSITE; PS00879; ODR DC 5.32; 1.
 DR PROSITE; PS00879; ODR DC 5.33; 1.
 DR PROSITE; PS00879; ODR DC 5.34; 1.
 DR PROSITE; PS00879; ODR DC 5.35; 1.
 DR PROSITE; PS00879; ODR DC 5.36; 1.
 DR PROSITE; PS00879; ODR DC 5.37; 1.
 DR PROSITE; PS00879; ODR DC 5.38; 1.
 DR PROSITE; PS00879; ODR DC 5.39; 1.
 DR PROSITE; PS00879; ODR DC 5.40; 1.
 DR PROSITE; PS00879; ODR DC 5.41; 1.
 DR PROSITE; PS00879; ODR DC 5.42; 1.
 DR PROSITE; PS00879; ODR DC 5.43; 1.
 DR PROSITE; PS00879; ODR DC 5.44; 1.
 DR PROSITE; PS00879; ODR DC 5.45; 1.
 DR PROSITE; PS00879; ODR DC 5.46; 1.
 DR PROSITE; PS00879; ODR DC 5.47; 1.
 DR PROSITE; PS00879; ODR DC 5.48; 1.
 DR PROSITE; PS00879; ODR DC 5.49; 1.
 DR PROSITE; PS00879; ODR DC 5.50; 1.
 DR PROSITE; PS00879; ODR DC 5.51; 1.
 DR PROSITE; PS00879; ODR DC 5.52; 1.
 DR PROSITE; PS00879; ODR DC 5.53; 1.
 DR PROSITE; PS00879; ODR DC 5.54; 1.
 DR PROSITE; PS00879; ODR DC 5.55; 1.
 DR PROSITE; PS00879; ODR DC 5.56; 1.
 DR PROSITE; PS00879; ODR DC 5.57; 1.
 DR PROSITE; PS00879; ODR DC 5.58; 1.
 DR PROSITE; PS00879; ODR DC 5.59; 1.
 DR PROSITE; PS00879; ODR DC 5.60; 1.
 DR PROSITE; PS00879; ODR DC 5.61; 1.
 DR PROSITE; PS00879; ODR DC 5.62; 1.
 DR PROSITE; PS00879; ODR DC 5.63; 1.
 DR PROSITE; PS00879; ODR DC 5.64; 1.
 DR PROSITE; PS00879; ODR DC 5.65; 1.
 DR PROSITE; PS00879; ODR DC 5.66; 1.
 DR PROSITE; PS00879; ODR DC 5.67; 1.
 DR PROSITE; PS00879; ODR DC 5.68; 1.
 DR PROSITE; PS00879; ODR DC 5.69; 1.
 DR PROSITE; PS00879; ODR DC 5.70; 1.
 DR PROSITE; PS00879; ODR DC 5.71; 1.
 DR PROSITE; PS00879; ODR DC 5.72; 1.
 DR PROSITE; PS00879; ODR DC 5.73; 1.
 DR PROSITE; PS00879; ODR DC 5.74; 1.
 DR PROSITE; PS00879; ODR DC 5.75; 1.
 DR PROSITE; PS00879; ODR DC 5.76; 1.
 DR PROSITE; PS00879; ODR DC 5.77; 1.
 DR PROSITE; PS00879; ODR DC 5.78; 1.
 DR PROSITE; PS00879; ODR DC 5.79; 1.
 DR PROSITE; PS00879; ODR DC 5.80; 1.
 DR PROSITE; PS00879; ODR DC 5.81; 1.
 DR PROSITE; PS00879; ODR DC 5.82; 1.
 DR PROSITE; PS00879; ODR DC 5.83; 1.
 DR PROSITE; PS00879; ODR DC 5.84; 1.
 DR PROSITE; PS00879; ODR DC 5.85; 1.
 DR PROSITE; PS00879; ODR DC 5.86; 1.
 DR PROSITE; PS00879; ODR DC 5.87; 1.
 DR PROSITE; PS00879; ODR DC 5.88; 1.
 DR PROSITE; PS00879; ODR DC 5.89; 1.
 DR PROSITE; PS00879; ODR DC 5.90; 1.
 DR PROSITE; PS00879; ODR DC 5.91; 1.
 DR PROSITE; PS00879; ODR DC 5.92; 1.
 DR PROSITE; PS00879; ODR DC 5.93; 1.
 DR PROSITE; PS00879; ODR DC 5.94; 1.
 DR PROSITE; PS00879; ODR DC 5.95; 1.
 DR PROSITE; PS00879; ODR DC 5.9

```
RT host specificities.;
RL Nature 417:459-463(2002).
DR EMBL; AE011966; AAM38051.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 596 AA; 64542 MW; 40B880665884FCD7 CRC64;

Query Match 46.0%; Score 45.5; DB 16; Length 596;
Best Local Similarity 60.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 1; Mismatches 6; Indels 1; Gaps 1;

QY 1 LQAVRLALSGSDGRVGS 20
   |||||
DB 27 LQAAARSA-SGRAGKGD 45

RESULT 16
Q9XDG7
ID Q9XDG7 PRELIMINARY; PRT; 139 AA.
AC Q9XDG7;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Afad.
DE Afad.
GN Afad.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=239 KH 89;
RX MEDLINE=99426795; PubMed=10496877;
RA Lalioui L., Jouve M., Gounon P., Le Bouguenec C.;
RT "Molecular cloning and characterization of the afa-7 and afa-8 gene
RT clusters encoding afimbrial adhesins in Escherichia coli strains
RT associated with diarrhea or septicemia in calves.";
RL Infect. Immun. 67:5048-5059(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=239 KH 89;
RX MEDLINE=21101044; PubMed=11159989;
RA Lalioui L., Le Bouguenec C.C.;
RT "afa-8 Gene cluster is carried by a pathogenicity island inserted into
RT the tRNA(Phe) of human and bovine pathogenic Escherichia coli
RT isolates.";
RL Infect. Immun. 69:937-948(2001).
DR EMBL; AF072900; RAD44023.1; -.
DR InterPro; IPR008394; Afad.
DR Pfam; PF05775; Afad; 1.
SQ SEQUENCE 139 AA; 15028 MW; C23D5C96B176831C CRC64;

Query Match 45.5%; Score 45; DB 2; Length 139;
Best Local Similarity 58.8%; Pred. No. 38;
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 LQAVRLALSGSDGRV 17
   :|||:|:|:|:|
DB 19 LAQAVELSLTSDGRS 35

RESULT 17
Q851F8
ID Q851F8 PRELIMINARY; PRT; 271 AA.
AC Q851F8;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical protein.
GN OSJNB0011H13.4.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
```

```
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Buell C.R., Yuan Q., Ouyang S., Liu J., Ganeberg K., Jones K.M.,
RA Overton II L.L., Tsitrin T., Kim M.M., Bera J.J., Jin S.S.,
RA Padrosh D.W., Tallon L.J., Koo H., Zismann V., Hsiao J., Blunt S.,
RA Vanaken S.S., Riedmuller S.B., Utterback T.T., Feidblyum T.V.,
RA Yang Q.Q., Haas B.J., Suh B.B., Peterson J.J., Quackenbush J.,
RA White O., Salzberg S.L., Fraser C.M.;
RT "Oryza sativa chromosome 3 BAC OSJNB0011H13 genomic sequence.";
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC120983; AAO38479.1; -.
KW Hypothetical protein.
SQ SEQUENCE 271 AA; 28326 MW; 394CB729A68B9B7B CRC64;

Query Match 45.5%; Score 45; DB 10; Length 271;
Best Local Similarity 58.8%; Pred. No. 78;
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 ORAVRLALSGSDGRVGS 18
   :|||:|:|:|:|
DB 49 RRRRLALSGADARGA 65

RESULT 18
Q910G7
ID Q910G7 PRELIMINARY; PRT; 359 AA.
AC Q910G7;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein PA2671.
GN PA2671.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saler M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004695; AAG06059.1; -.
DR PIR; E83312; E83312.
DR GO; GO:0005351; F:sugar porter activity; IEA.
DR GO; GO:0009401; P:phosphoenolpyruvate-dependent sugar phospho. .; IEA.
DR InterPro; IPR002114; HPr Serp S.
DR PROSITE; PS00589; PTS_HPR_SER; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 359 AA; 39404 MW; ACAD0EFA9187B566 CRC64;

Query Match 45.5%; Score 45; DB 16; Length 359;
Best Local Similarity 62.5%; Pred. No. 1.1e+02;
Matches 10; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 LQAVRLALSGSDGRV 16
   |||:|:|:|:|
DB 126 LDRSARLSLSDGRV 141

RESULT 19
Q8BR66
ID Q8BR66 PRELIMINARY; PRT; 692 AA.
AC Q8BR66;
```

```
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Plenty of SH3 domains (Fragment).
GN POSH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Brain;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK045470; BAC32385.1; -.
DR MGD; MGI:1913066; Posh.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 3.
DR PRINTS; PR00452; SH3DOMAIN.
DR ProDom; PD000066; SH3; 3.
DR SMART; SM00326; SH3; 3.
DR PROSITE; PS50002; SH3; 3.
FT NON_TER 1
SQ SEQUENCE 692 AA; 71660 MW; 152FC62B3BF1920D CRC64;

Query Match 45.5%; Score 45; DB 11; Length 692;
Best Local Similarity 65.0%; Pred. No. 2.1e+02;
Matches 13; Conservative 0; Mismatches 5; Indels 2; Gaps 1;

QY 1 LQRAV--RLALSGSDGRVGS 18
DB 559 LQAVGPPELPLGGSHGRVGS 578

RESULT 20
Q22354
ID Q22354 PRELIMINARY; PRT; 862 AA.
AC Q22354;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Hypothetical protein T08H4.1 in chromosome II.
GN T08H4.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Stellyes L.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U39470; AAC71118.1; -.
DR PIR; T34342; T34342.
DR WormPep; T08H4.1; CE04926.
DR InterPro; IPR001849; PH.
DR InterPro; IPR000219; RhoGEF.
DR Pfam; PF0621; RhoGEF; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00325; RhoGEF; 1.
DR PROSITE; PS50010; DH_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 862 AA; 97540 MW; 54CD47C47D5ADD28 CRC64;

Query Match 45.5%; Score 45; DB 5; Length 862;
Best Local Similarity 53.3%; Pred. No. 2.7e+02;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 6 RLALSGSDGRVSGC 20
```

```
Db 710 RFLNGSGGAVGSSC 724
|:|:| | ||| |
RESULT 21
O70254
ID O70254 PRELIMINARY; PRT; 892 AA.
AC O70254;
DT 01-AUG-1998 (TReMBLrel. 07, Created)
DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Plenty of SH3s.
GN POSH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98151363; PubMed=9482736;
RA Tapon N., Nagata K., Lamarche N., Hall A.;
RT "A new rac target POSH is an SH3-containing scaffold protein involved
RT in the JNK and NF-kappaB signalling pathways."
RL EMBO J. 17:1395-1404(1998).
CC -!- SIMILARITY: CONTAINS 4 SH3 DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR EMBL; AF030131; AAC40070.1; -.
DR PIR; T09071; T09071.
DR HSSP; P41240; ICSK.
DR MGD; MGI:1913066; Posh.
DR InterPro; IPR000108; Neu_cyt_fact_2.
DR InterPro; IPR001452; SH3.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF00018; SH3; 4.
DR PRINTS; PR00499; P67PHOX.
DR ProDom; PD000066; SH3; 4.
DR SMART; SM00184; RING; 1.
DR SMART; SM00326; SH3; 4.
DR PROSITE; PS50002; SH3; 4.
DR PROSITE; PS50018; ZF_RING_1; 1.
DR PROSITE; PS50089; ZF_RING_2; 1.
KW Metal-binding; SH3 domain; Zinc; Zinc-finger.
SQ SEQUENCE 892 AA; 93435 MW; 5A103085777B3B0 CRC64;

Query Match 45.5%; Score 45; DB 11; Length 892;
Best Local Similarity 65.0%; Pred. No. 2.8e+02;
Matches 13; Conservative 0; Mismatches 5; Indels 2; Gaps 1;

QY 1 LQRAV--RLALSGSDGRVGS 18
DB 759 LQAVGPPELPLGGSHGRVGS 778

RESULT 22
O44319
ID O44319 PRELIMINARY; PRT; 1151 AA.
AC O44319;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Reverse transcriptase (Fragment).
OS Anurida maritima.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Collembola; Arthropoleona;
OC Poduroidea; Neanuridae; Pseudochorutinae; Anurida.
OX NCBI_TaxID=64695;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98157515; PubMed=9515960;
RA Burke W.D., Malik H.S., Lathe W.C. III, Eichbush T.H.;
RT "Are retrotransposons long-term hitchhikers?";
RL Nature 392:141-142(1998).
```

[2]
RN SEQUENCE FROM N.A.
RX MEDLINE=99261661; PubMed=10331276;
RA Burke W.D., Malik H.S., Jones J.P., Eickbush T.H.;
RT "The domain structure and retrotransposition mechanism of R2 elements
are conserved throughout arthropods.";
RL Mol. Biol. Evol. 16:502-511(1999).
DR EMBL; AF015815; AAC34903.1; -;
DR FIC; T30936; T30936.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
DR GO; GO:0016740; F:transfrase activity; IEA.
DR GO; GO:0006278; P:RNA dependent DNA replication; IEA.
DR InterPro; IPR000477; RVTe.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00078; rvt; 1.
DR SMART; SM00355; Znf_C2H2; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 1.
KW Metal-binding; RNA-directed DNA polymerase; Transferase; Zinc;
FT Zinc-finger.
FT NON TER
SQ SEQUENCE 1151 AA; 127630 MW; 0566B15965585B15 CRC64;
Query Match 45.5%; Score 45; DB 5; Length 1151;
Best Local Similarity 64.3%; Pred. No. 3.6e+02;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 1 LQRAVRLALSGSDG 14
DB 804 LDRMVRVALTGKDG 817
RESULT 23
ID Q826G9 PRELIMINARY; PRT; 1399 AA.
AC Q826G9;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN SAV7224.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
avermitilis: deducing the ability of producing secondary
metabolites.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=WA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
microorganism Streptomyces avermitilis.";
RL Nat. Biotechnol. 21:526-531(2003).
DR EMBL; AF005049; BAC74935.1; -;
DR GO; GO:0005743; C:mitochondrial inner membrane; IEA.
DR GO; GO:0005488; F:binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR001993; Mitoch_cARRIER.
DR PROSITE; PS00215; MITOCH_CARRIER; 1.
KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 1399 AA; 153517 MW; 7C4786D3EBE6417A5 CRC64;
Query Match 45.5%; Score 45; DB 16; Length 1399;
Best Local Similarity 55.6%; Pred. No. 4.5e+02;
Matches 10; Conservative 6; Mismatches 6; Indels 0; Gaps 0;
QY 1 LQRAVRLALSGSDG 18
DB 901 LRRVRVALSADGHRGS 918
RESULT 24
ID Q9VLR3 PRELIMINARY; PRT; 158 AA.
AC Q9VLR3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CG8360 protein.
GN CG8360.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berhan B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mileshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrtk R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AR003620; AAF52621.1; -;
DR HSSP; P13652; 1CTT.
DR FlyBase; FBgn0032001; CG8360.
DR GO; GO:0004126; F:Cytidine deaminase activity; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.

```

DR GO: 0046087; P:cytidine metabolism; IEA.
DR InterPro: IPR006262; Cyt deam tetra.
DR Pfam: PF00383; dCMP_cyt_deam.
DR TIGRFAMs: TIGR01354; cyt_deam_tetra.
SQ SEQUENCE 158 AA; 16999 MW; 30470D22168301A9 CRC64;

Query Match      44.4%; Score 44; DB 5; Length 158;
Best Local Similarity 58.8%; Pred. No. 63;
Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 4 AVRLALSGSDGRVSGC 20
DB 47 AVGAALRTSDGTIYSGC 63

RESULT 25
QBSUF1
ID QBSUF1 PRELIMINARY; PRT; 256 AA.
AC QBSUF1;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE Hypothetical protein ECU10_0790.
GN ECU10 0790.
OS Encephalitozoon cuniculi.
OC Eukaryota; Fungi; Microsporidia; Unikaryonidae; Encephalitozoon.
OX NCBI_TaxID=6035;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GB-M1;
RA Genoscope;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=GB-M1;
RX MEDLINE=21576510; PubMed=11719806;
RA Katinka M.D., Duprat S., Cornillot E., Metenier G., Thomarat F.,
RA Preiner G., Barbe V., Peyretailade E., Brottier P., Wincker P.,
RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,
RA Weissenbach J., Vivares C.P.;
RT "Genome sequence and gene compaction of the eukaryote parasite
RT Encephalitozoon cuniculi."
RL Nature 414:450-453(2001).
DR EMBL; AL590449; CAD25798.1; -.
XW Hypothetical protein.
SQ SEQUENCE 256 AA; 29644 MW; F1F09D015BC0462E CRC64;

Query Match      44.4%; Score 44; DB 5; Length 256;
Best Local Similarity 56.2%; Pred. No. 1.1e+02;
Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 LQAVRLALSGSDGRV 16
DB 17 LLRSTRATISGEDGRV 32

RESULT 26
QBAVP9
ID QBAVP9 PRELIMINARY; PRT; 263 AA.
AC QBAVP9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to heterogenous nuclear ribonucleoprotein C (C1/C2).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;

```

```

RA Klein S., Strausberg R.;
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC041534; AAH41534.1; -.
DR GO: 0030529; C:ribonucleoprotein complex; IEA.
DR GO: 0039013; C:viral nucleocapsid; IEA.
DR GO: 0000367; F:nucleic acid binding; IEA.
DR InterPro: IPR00504; RNA_rec_mot.
DR Pfam: PF00076; rrm; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS0102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
KW Nucleocapsid; Ribonucleoprotein.
SQ SEQUENCE 263 AA; 29195 MW; 6049D06BC031C479 CRC64;

Query Match      44.4%; Score 44; DB 13; Length 263;
Best Local Similarity 44.4%; Pred. No. 1.1e+02;
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 2 QRAVRLALSGSDGRVSG 19
DB 61 ERTARTAVAGEDGRWAG 78

RESULT 27
QBUEX7
ID QBUEX7 PRELIMINARY; PRT; 296 AA.
AC QBUEX7;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE NAD dependent epimerase/dehydratase family protein.
GN ATU1626 OR AGR C 3001.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eissen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58."
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmlel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RL Science 294:2323-2328(2001).
DR EMBL; AF009120; AAL42628.1; -.
DR EMBL; AF008085; AAK87405.1; -.
DR PIR; AF2776; AF2776.
DR PIR; D97556; D97556.
KW Complete proteome.
SQ SEQUENCE 296 AA; 31988 MW; 6D24DA3D89C282E3 CRC64;

Query Match      44.4%; Score 44; DB 16; Length 296;
Best Local Similarity 61.5%; Pred. No. 1.2e+02;

```



```

Matches      8;  Conservative      3;  Mismatches      2;  Indels      0;  Gaps      0;

QY      5 VRLLALSGSGRVG 17
      :|:|:|:|:|
Db      1 MRVLTGSSGRVG 13

RESULT 28
Q9RYP2      PRELIMINARY;      PRT;      376 AA.
AC      Q9RYP2;
DT      01-WAY-2000 (TREMBlrel. 13, Created)
DT      01-WAY-2000 (TREMBlrel. 13, Last sequence update)
DT      01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE      Adenine deaminase-related protein.
GN      DRA0268.
OS      Deinococcus radiodurans.
OC      Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC      Deinococcaceae; Deinococcus.
OX      NCBI_TaxID=1299;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
RX      MEDLINE=20036896; PubMed=10567266;
RA      White O., Eissen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA      Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA      Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA      Vamathevan J.J., Lam P., McDonnell L., Utterback T., Zalewski C.,
RA      Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA      Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA      Fraser C.M.;
RT      "Genome sequence of the radioresistant bacterium Deinococcus
RT      radiodurans R1.";
RL      Science 286:1571-1577(1999).
DR      EMBL; AB001863; AAP12376.1; -.
DR      PIR; C75580; C75580.
DR      TIGR; DRA0268; -.
KW      Complete proteome.
SQ      SEQUENCE 376 AA; 39845 MW; 7AB7FF32F8C45651 CRC64;

Query Match      44.4%; Score 44; DB 16; Length 376;
Best Local Similarity 64.7%; Pred. No. 1.6e+02;
Matches 11; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY      3 RAVRLALSGSGRVGSG 19
      ||| ||| ||| |||
Db      275 RAVAPALGSDRPPASG 291

RESULT 29
Q9A9Y9      PRELIMINARY;      PRT;      478 AA.
AC      Q9A9Y9;
DT      01-JUN-2001 (TREMBlrel. 17, Created)
DT      01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT      01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE      Aldehyde dehydrogenase.
GN      CC0822.
OS      Caulobacter crescentus.
OC      Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
OC      Caulobacteraceae; Caulobacter.
OX      NCBI_TaxID=155892;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=ATCC 19089 / CB15;
RX      MEDLINE=21173698; PubMed=11259647;
RA      Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.B.,
RA      Eissen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA      Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA      DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA      Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA      Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA      Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;

```

```

RT      "Complete genome sequence of Caulobacter crescentus.";
RL      Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR      EMBL; AB005759; AAK22807.1; -.
DR      PIR; C87351; C87351.
DR      HSSP; P05091; 1CW3.
DR      TIGR; CC0822; -.
DR      GO; GO:0016491; F:oxidoreductase activity; IEA.
DR      GO; GO:0008152; P:metabolism; IEA.
DR      InterPro; IPR002086; Aldehyde_dehydr.
DR      Pfam; PF00171; aldedh; 1.
DR      PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; 1.
DR      PROSITE; PS00687; ALDEHYDE_DEHYDR_GLU; 1.
KW      Complete proteome.
SQ      SEQUENCE 478 AA; 50576 MW; 04C87F388445C3A9 CRC64;

Query Match      44.4%; Score 44; DB 16; Length 478;
Best Local Similarity 50.0%; Pred. No. 2e+02;
Matches 10; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY      1 LQRAVRLALSGSGRVGSGC 20
      :|:|:|:|:|
Db      262 LERAVAILDGSFFATGQRC 281

RESULT 30
Q8XW48      PRELIMINARY;      PRT;      514 AA.
AC      Q8XW48;
DT      01-MAR-2002 (TREMBlrel. 20, Created)
DT      01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT      01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE      Probable polynucleotide adenyllyltransferase protein
DE      (EC 2.7.7.19).
GN      PCNB OR RSC2627 OR R900931.
OS      Ralstonia solanacearum (Pseudomonas solanacearum).
OC      Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC      Burkholderiaceae; Ralstonia.
OX      NCBI_TaxID=305;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=GM11000;
RX      MEDLINE=21681879; PubMed=11823852;
RA      Salanoubat M., Genin S., Artiguenave P., Gouzy J., Mangenot S.,
RA      Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA      Chandler M., Choigne N., Claudel-Renard C., Cunnac S., Demange N.,
RA      Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA      Signier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA      Weissenbach J., Boucher C.A.;
RT      "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL      Nature 415:497-502(2002).
DR      EMBL; AL646071; CAD16334.1; -.
DR      GO; GO:0004652; F:polynucleotide adenyllyltransferase activity; IEA.
DR      GO; GO:0003723; F:RNA binding; IEA.
DR      GO; GO:0016740; F:transferase activity; IEA.
DR      GO; GO:0006396; P:RNA processing; IEA.
DR      InterPro; IPR002646; PolyA_pol.
DR      Pfam; PF01743; PolyA_pol; 1.
KW      Transferase; Nucleotidyltransferase; Complete proteome.
SQ      SEQUENCE 514 AA; 57329 MW; 2189D3D0D438FDD5 CRC64;

Query Match      44.4%; Score 44; DB 16; Length 514;
Best Local Similarity 55.6%; Pred. No. 2.2e+02;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY      2 QRAVRLALSGSGRVGSG 19
      ||| ||| ||| |||
Db      .329 QRFVQLALDNTDERVKAG 346

Search completed: May 13, 2004, 06:55:18
Job time : 36.3333 secs

```

This Page Blank (uspto)